

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 17, 2004, 06:59:28 ; Search time 18.3365 Seconds  
(without alignments)  
1374.429 Million cell updates/sec

Title: US-10-057-951-2\_COPY\_2\_263  
Perfect score: 1388  
Sequence: 1 LLANVQARFLVSNMLLARYG.....PVDPQEGSTPLMGQAGTPGA 262

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : PIR 78: \*  
1: Pirl: \*  
2: Pirl2: \*  
3: Pirl3: \*  
4: Pirl4: \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	154.5	11.1	562	1 UKHUT	t-plasminogen acti
2	154	11.1	559	1 A35029	t-plasminogen acti
3	147.5	10.6	559	1 A29941	t-plasminogen acti
4	146.5	10.6	655	1 A45688	hepatocyte growth
5	145.5	10.5	291	2 I38098	t-plasminogen acti
6	142	10.2	810	1 PLHU	plasmin (EC 3.4.21
7	138	9.9	169	2 A40522	plasmin (EC 3.4.21
8	138	9.9	433	1 UKBAY	u-plasminogen acti
9	137	9.9	716	1 JC5061	macrophage-stimula
10	135.5	9.8	442	1 UKPG	u-plasminogen acti
11	134	9.7	431	2 JS0599	t-plasminogen acti
12	134	9.7	477	1 A34369	t-plasminogen acti
13	134	9.7	477	2 JS0598	t-plasminogen acti
14	134	9.7	716	1 A40332	macrophage-stimula
15	132	9.5	394	2 JS0600	t-plasminogen acti
16	132	9.5	431	1 UKHU	u-plasminogen acti
17	130	9.4	810	2 I46260	plasmin (EC 3.4.21
18	128.5	9.3	434	1 A35005	u-plasminogen acti
19	127	9.1	433	1 JS0560	u-plasminogen acti
20	126.5	9.0	728	1 A60185	hepatocyte growth
21	125.5	9.0	560	1 JC4795	plasma hyaluronan-
22	124.5	9.0	810	2 B30848	thrombin (EC 3.4.21
23	124	8.9	622	1 TBHO	thrombin (EC 3.4.21
24	124	8.9	728	1 JH0579	hepatocyte growth
25	124	8.9	812	1 PLBO	plasmin (EC 3.4.21
26	123	8.9	593	2 S45281	coagulation factor
27	122.5	8.8	4548	1 S06657	apoptosis(a) (EC
28	122	8.8	603	2 S28941	coagulation factor
29	121.5	8.8	728	1 A35644	hepatocyte growth

30	120	8.6	433	1 UKMS	u-plasminogen acti
31	120	8.6	790	1 PLRG	plasmin (EC 3.4.21
32	119.5	8.6	477	2 JS0597	t-plasminogen acti
33	119	8.6	710	1 IS1283	hepatocyte growth
34	118.5	8.5	615	1 KFHU12	coagulation factor
35	118	8.5	432	1 S18932	u-plasminogen acti
36	117	8.4	625	1 TBBO	thrombin (EC 3.4.21
37	116	8.4	1420	2 A32869	apolipoprotein(a)
38	115.5	8.3	711	1 A47136	macrophage-stimula
39	114.5	8.2	812	1 FLMS	plasmin (EC 3.4.21
40	114	8.2	618	2 A35827	thrombin (EC 3.4.21
41	113	8.1	558	2 JC5878	plasma hyaluronan-
42	112	8.0	455	1 A61545	plasmin (EC 3.4.21
43	110.5	8.0	685	1 A48289	neurotrophic recep
44	108.5	7.8	123	2 C61545	plasmin (EC 3.4.21
45	107	7.7	617	2 S10511	thrombin (EC 3.4.21

ALIGNMENTS

RESULT 1

UKHUT  
t-plasminogen activator (EC 3.4.21.68) precursor [validated] - human  
N;Alternate names: t-PA; tissue plasminogen activator  
C;Species: Homo sapiens (man)  
C;Date: 14-Nov-1983 #sequence revision 14-Nov-1983 #text change 08-Dec-2000  
C;Accession: A94004; A23529; JT0562; A93293; S02125; A91343; A93951; A91322; A54645; I6  
R;NV: T.; Elgh, F.; Lund, B.  
Proc. Natl. Acad. Sci. U.S.A. 81, 5355-5359, 1984  
A;Title: The structure of the human tissue-type plasminogen activator gene: correlation  
A;Reference number: A94004; MUID:84298137; PMID:6089198  
A;Accession: A94004  
A;Molecule type: DNA  
A;Residues: 1-562 <NYT>  
A;Cross-references: GB:I00141  
A;Note: the codon given for residue 93 (ACC) is inconsistent with the authors' translat  
R;Friesner Degen, S.J.; Rajput, B.; Reich, E.  
J. Biol. Chem. 261, 6972-6985, 1986  
A;Title: The human tissue plasminogen activator gene.  
A;Reference number: A23529; MUID:86196143; PMID:3009482  
A;Accession: A23529  
A;Molecule type: DNA  
A;Residues: 1-562 <DEG>  
A;Cross-references: GB:X03021; NID:g339817; PIDN:AAA98809.1; PID:g339818  
R;Itagaki, Y.; Yasuda, H.; Morinaga, T.; Mitsuda, S.; Higashio, K.  
Agric. Biol. Chem. 55, 1225-1232, 1991  
A;Title: Purification and characterization of tissue plasminogen activator secreted by  
A;Reference number: JT0562; MUID:91291340; PMID:1368681  
A;Accession: JT0562  
A;Molecule type: mRNA  
A;Residues: 31-562 <ITA>  
A;Cross-references: DBU:D01096; NID:g220128; PIDN:BAA00881.1; PID:g441174  
A;Experimental source: embryonic lung fibroblast IMR-90 cells  
A;Note: part of this sequence, including the amino end of the mature protein, was confi  
R;Pennica, D.; Holmes, W.E.; Kohr, W.J.; Harkins, R.N.; Vehar, G.A.; Ward, C.A.; Bennet  
Nature 301, 214-221, 1993  
A;Title: Cloning and expression of human tissue-type plasminogen activator cDNA in Esch  
A;Reference number: A93293; MUID:83115262; PMID:6337343  
A;Accession: A93293  
A;Molecule type: mRNA  
A;Residues: 1-562 <PEN>  
A;Cross-references: GB:I00141  
R;Sasaki, H.; Saito, Y.; Hayaishi, M.; Otsuka, K.; Niwa, M.  
Nucleic Acids Res. 16, 5695, 1988  
A;Title: Nucleotide sequence of the tissue-type plasminogen activator cDNA from human f  
A;Reference number: S02125; MUID:88262579; PMID:3133640  
A;Accession: S02125  
A;Status: translation not shown  
A;Molecule type: mRNA  
A;Residues: 1-562 <SAS>  
A;Cross-references: EMBL:X07393; NID:g37243; PIDN:CAA30302.1; PID:g37244



F;309-559/Product: t-plasminogen activator chain B #status predicted <BCH>  
 F;309-553/Domain: trypsin homology <TRY>  
 F;38-66,66-75,83-94,88-105,107-116,124-205,145-187,176-200,213-294,234-276,265-289,297-4  
 F;149,481/Binding site: carbonylde (Asn) (covalent) #status predicted  
 F;308-309/Cleavage site: Arg-ile (plasmin, trypsin) #status predicted  
 F;355,404,510/Active site: His, Asp, Ser #status predicted

Query Match 11.1%; Score 154; DB 1; Length 559;  
 Best Local Similarity 32.4%; Pred. No. 3.2e-05;  
 Matches 45; Conservative 14; Mismatches 54; Indels 26; Gaps 6;

QY 24 CFWDNGHLYREDQTSAPGLRCLNWLDAQGLASAPVS-----GAGNHSYCRNPDE 74  
 Db 124 CFEGQITRGVTSWTAENGACINW--NSALSQKYSARRNAIKLGLGNHNYCRNPDR 181

QY 75 DPGPWCYVSGEAGVPEKRCEDLRCE-----TTSQLPAFTTEQEAEGFG 123  
 Db 182 DVK-PWCYVF-KAGKYTFECSPACPKGPTDCYGVKGVYTRGTHFTT--SKASCLPW 237

QY 124 ADEVQVFPANALPASSEA 142  
 Db 238 NSMILIGKVTAWRANSOA 256

RESULT 3  
 A29941  
 t-Plasminogen activator (EC 3.4.21.68) precursor - mouse  
 C;Species: Mus musculus (house mouse)  
 C;Date: 10-Sep-1999 #sequence\_revision 10-Sep-1999 #text\_change 10-Sep-1999  
 C;Accession: A29941; S48205; S48207; S48206  
 R;Rickles, R.J.; Darrow, A.L.; Strickland, S.  
 J. Biol. Chem. 263, 1563-1569, 1988  
 A;Title: Molecular cloning of complementary DNA to mouse tissue plasminogen activator mP  
 A;Reference number: A29941; MUID:98087303; PMID:2826484  
 A;Accession: A29941  
 A;Molecule type: mRNA  
 A;Residues: 1-559 <RIC>  
 A;Cross-references: GB:J03520; NID:G202109; PIDN:AAA40470.1; PID:G202110  
 R;Lijnen, H.R.; van Hoef, B.; Beelen, V.; Collen, D.  
 Eur. J. Biochem. 224, 863-871, 1994  
 A;Title: Characterization of the murine plasma fibrinolytic system.  
 A;Reference number: S48202; MUID:95010076; PMID:7523120  
 A;Accession: S48205  
 A;Molecule type: protein  
 A;Residues: 33-37, 'X', 39-40 <LIJ>  
 A;Accession: S48207  
 A;Molecule type: protein  
 A;Residues: 309-316 <LIJ>  
 A;Accession: S48206  
 A;Molecule type: protein  
 A;Residues: 33-37, 'X', 39-40 <LIJ>  
 C;Superfamily: tissue plasminogen activator; EGF homology; fibronectin type I repeat hom  
 C;Keywords: fibrinolysis; glycoprotein; hydrolase; kringle; serine proteinase  
 F;1-17/Domain: signal sequence #status predicted <SIG>  
 F;18-29/Domain: propeptide #status predicted <PRO>  
 F;30-59/Product: t-plasminogen activator #status predicted <MAT>  
 F;30-308/Product: t-plasminogen activator chain A #status predicted <ACH>  
 F;38-75/Domain: fibronectin type I repeat homology <IF1>  
 F;83-116/Domain: EGF homology <EGF>  
 F;124-205/Domain: kringle homology <KR1>  
 F;213-294/Domain: kringle homology <KR2>  
 F;309-559/Product: t-plasminogen activator chain B #status predicted <BCH>  
 F;309-553/Domain: trypsin homology <TRY>  
 F;38-66,66-75,83-94,88-105,107-116,124-205,145-187,176-200,213-294,234-276,265-289,297-4  
 F;149,481/Binding site: carbonylde (Asn) (covalent) #status predicted  
 F;308-309/Cleavage site: Arg-ile (plasmin, trypsin) #status predicted  
 F;355,404,510/Active site: His, Asp, Ser #status predicted

Query Match 10.6%; Score 147.5; DB 1; Length 559;  
 Best Local Similarity 37.0%; Pred. No. 0.00011;  
 Matches 34; Conservative 11; Mismatches 34; Indels 13; Gaps 4;

QY 24 CFWDNGHLYREDQTSAPGLRCLNWLDAQGLASAPVS-----GAGNHSYCRNPDE 74

Db 124 CFEGQITRGVTSWTAENGACINW--NSSVLSIKPVNARRPNAIKLGLGNHNYCRNPDR 181  
 QY 75 DPGPWCYVSGEAGVPEKRCEDLRCEFTTSQ 106  
 Db 182 DLK-PWCYVF-KAGKYTFECSPACPKGKSE 211

RESULT 4  
 A46688  
 hepatocyte growth factor activator (EC 3.4.21.-) precursor [validated] - human  
 C;Species: Homo sapiens (man)  
 C;Date: 21-Sep-1993 #sequence\_revision 25-Aug-1995 #text\_change 08-Dec-2000  
 C;Accession: A46688  
 R;Miyazawa, K.; Shimomura, T.; Kitamura, A.; Kondo, J.; Morimoto, Y.; Kitamura, N.  
 J. Biol. Chem. 268, 10024-10028, 1993  
 A;Title: Molecular cloning and sequence analysis of the cDNA for a human serine protease  
 d coagulation factor XII.  
 A;Reference number: A46688; MUID:93252878; PMID:7683665  
 A;Accession: A46688  
 A;Molecule type: mRNA  
 A;Residues: 1-655 <MIY>  
 A;Cross-references: DDBJ:D14012; NID:G219680; PIDN:BAA03113.1; PID:G219681  
 A;Experimental source: liver (mRNA); serum (protein)  
 A;Note: sequence extracted from NCBI backbone (NCBIN:131227, NCBIIP:131228)  
 A;Note: parts of the sequence, including the amino ends of the heavy and light chains,  
 C;Genetics:  
 A;Gene: GDB:HGPAC; HGFA; HGPAP  
 A;Cross-references: GDB:9954514  
 A;Map position: 4p16-4p16  
 C;Function:  
 A;Description: activates hepatocyte growth factor by specific proteolytic cleavage  
 A;Pathway: tissue repair and regeneration  
 C;Superfamily: coagulation factor XII; EGF homology; fibronectin type I repeat homology  
 C;Keywords: glycoprotein; hydrolase; kringle; liver; plasma; serine proteinase  
 F;1-34/Domain: signal sequence #status predicted <SIG>  
 F;108-148/Domain: fibronectin type II repeat homology <IF2>  
 F;164-197/Domain: EGF homology <EG1>  
 F;202-237/Domain: fibronectin type I repeat homology <IF1>  
 F;245-278/Domain: EGF homology <EG2>  
 F;286-367/Domain: kringle homology <KRG>  
 F;373-407/Product: hepatocyte growth factor activator light chain #status experimental  
 F;408-655/Product: hepatocyte growth factor activator heavy chain #status experimental  
 F;408-641/Domain: trypsin homology <TRY>  
 F;40,48,290,468,492,546/Binding site: carbonylde (Asn) (covalent) #status predicted  
 F;164-175,169-186,188-197,202-230,228-237,245-256,250-267,269-278,286-367,307-349,338-  
 F;447,497,598/Active site: His, Asp, Ser #status predicted

Query Match 10.6%; Score 146.5; DB 1; Length 655;  
 Best Local Similarity 36.9%; Pred. No. 0.00017;  
 Matches 41; Conservative 10; Mismatches 41; Indels 19; Gaps 6;

QY 24 CFWDNGHLYREDQTSAPGLRCLNWLDAQGLASAPVS-----LDAQGLASAPVSGAGNHSYCRNPDED 75  
 Db 286 CFLNGTGYRCAVSTASGLSCLAWNSDLYQLHVDV-VGMAALLGLGHPAYCRNPDDND 344

QY 76 PRGPWCYVSGEAGV-----EKRPCEDLRCEFTTSQALPAFTTIOQE-ASEG 121  
 Db 345 ER-PWCYVVKDSALSWEVCLEACESL---TRVLSPDLLATLPPEPASPG 390

RESULT 5  
 I38098  
 t-plasminogen activator precursor, inactive endothelial splice form - human  
 N;Alternate names: tissue plasminogen activator  
 C;Species: Homo sapiens (man)  
 C;Date: 17-May-1996 #sequence\_revision 17-May-1996 #text\_change 22-Jun-1999  
 C;Accession: I38098; S01678  
 R;Siebert, P.D.; Fong, K.  
 Nucleic Acids Res. 18, 1086, 1990  
 A;Title: Variant tissue type plasminogen activator (PLAT) cDNA obtained from human end  
 A;Reference number: I38098; MUID:90192128; PMID:1969145  
 A;Accession: I38098





A: Cross-references: GDB:119498; OXIM:173350  
A: Map position: 6q26-6q27  
A: Introns: 17/1; 62/2; 98/1; 136/2; 183/1; 223/2; 263/1; 317/2; 366/1; 419/2; 480/1; 512/2  
C: Function:  
A: Description: dissolves the fibrin of blood clots; acts as a proteolytic factor in a variety of cells; acts as a proteolytic factor in the walls of the Graafian follicle; also activates the urokinase-type plasminogen activator; also activates the urokinase-type plasminogen activator; also activates the urokinase-type plasminogen activator  
A: Pathway: fibrinolysis  
C: Superfamily: plasmin; kringle homology; plasminogen-related protein precursor homology  
C: Keywords: angiotensin inhibitor; blood; duplication; fibrinolysis; glycoprotein; hydrolase; plasminogen-related protein precursor homology  
F: 1-96/Domain: plasminogen-related protein precursor homology  
F: 1-19/Domain: signal sequence #status predicted <SIG>  
F: 19-810/Product: plasminogen #status experimental <PRO>  
F: 20-96/Domain: activation peptide #status experimental <APT>  
F: 79-466/Product: angiotensin #status experimental <AST>  
F: 97-580.581-810/Product: plasmin #status experimental <MAT>  
F: 97-580/Domain: plasmin chain A #status experimental <CHA>  
F: 103-181/Domain: kringle homology <KR1>  
F: 185-262/Domain: kringle homology <KR2>  
F: 275-352/Domain: kringle homology <KR3>  
F: 377-454/Domain: kringle homology <KR4>  
F: 481-560/Domain: kringle homology <KR5>  
F: 550-580.581-810/Product: microplasmin #status experimental <MMT>

Query Match 10.2%; Score 142; DB 1; Length 810;  
Best Local Similarity 37.2%; Pred. No. 0.00051;  
Matches 32; Conservative 8; Mismatches 34; Indels 12; Gaps 4;

QY 24 CFWDNGHLYREDOTSPAPGLRCLNWIDA-----QSGLASAPVSGAGNHSCRNPDDEPRG 78  
DB 103 CRTGNNGKVRGTWSTKXNGITCKWSTSPHRRFRFSPATHPSEGL-EENYCRNPNDPQG 161  
QY 79 PWCYVSGEAGVPEKR--PCEDLRCP 102  
DB 162 PWCYTTD---PEKRYDYCDLCEE 183

RESULT 7  
A40522  
plasmin (EC 3.4.21.7) precursor - rat (fragment)  
C: Species: Rattus norvegicus (Norway rat)  
C: Date: 28-Feb-1992 #sequence\_revision 17-Apr-1993 #text\_change 16-Jul-1999  
C: Accession: A40522  
R: Kanalas, J.J.; Makker, S.P.  
J. Biol. Chem. 266, 10825-10829, 1991  
A: Title: Identification of the rat Heymann nephritis autoantigen (GP330) as a receptor  
A: Reference number: A40522; MOID: I91250378; PMID: 1645711  
A: Accession: A40522  
A: Status: preliminary  
A: Molecule type: mRNA  
A: Residues: 1-169 <RAN>  
A: Cross-references: GB:M62832; NID:G206215; PIDN:AAA41884.1; PID:G554488  
A: Note: the authors translated the codon TCT for residue 76 as Ala  
C: Superfamily: plasmin; kringle homology; plasminogen-related protein precursor homology  
C: Keywords: fibrinolysis; glycoprotein; hydrolase; kringle; serine proteinase  
F: 34-112/Domain: kringle homology <RG>  
F: 34-112.55-95.83-107/Disulfide bonds: #status predicted

Query Match 9.9%; Score 138; DB 2; Length 169;  
Best Local Similarity 31.8%; Pred. No. 0.00018;  
Matches 35; Conservative 13; Mismatches 48; Indels 14; Gaps 5;

QY 24 CFWDNGHLYREDOTSPAPGLRCLNWIDA-----QSGLASAPVSGAGNHSCRNPDDEPRG 78  
DB 34 CYQNGKSYRGTSSTNTGKKCSWYMTFPHSHSKTPANFPDPSGL-EMNYCRNPNDPQG 92  
QY 79 PWCYVSGEAGVPEKR--PCEDLRCPPTTSQALPAFTTEIOEASGPGADE 126  
DB 93 PWCFTTD---FSVRVEYCNLKECSITGGGV--AESAIVPQVPSAQTSE 136

RESULT 8  
UKGAY  
u-plasminogen activator (EC 3.4.21.73) precursor - yellow baboon

C;Species: Papio cynocephalus, Papio hamadryas cynocephalus (yellow baboon)  
C;Date: 31-Dec-1991 #sequence\_revision 31-Dec-1991 #text\_change 18-Jun-1999  
C;Accession: S14687; S08651  
R;Au, Y. P. T.; Wang, T. W.; Clowes, A. W.  
Nucleic Acids Res. 18, 3411, 1990  
A;Title: Nucleotide and deduced amino acid sequences of baboon urokinase-type plasminogen activator  
A;Reference number: S14687; MUID: 90287734; PMID: 2113276  
A;Accession: S14687  
A;Molecule type: mRNA  
A;Residues: 1-433 <AUT>  
A;Cross-references: EMBL:X51935; NID:G381130; PIDN:CAA36200.1; PID:G38131  
C;Superfamily: urokinase-type plasminogen activator; EGF homology; kringle homology; try  
C;Keywords: glycoprotein; heterodimer; hydrolase; kringle; serine proteinase  
F;1-20/Domain: signal sequence #status predicted <SIG>  
F;21-176/Product: plasminogen activator chain A #status predicted <ACH>  
F;30-61/Domain: EGF homology <EGF>  
F;69-150/Domain: kringle homology <KRG>  
F;178-433/Product: plasminogen activator chain B #status predicted <BCH>  
F;178-421/Domain: trypsin homology <TRY>  
F;167-298, 208-224, 216-287, 315-394, 347-363, 374-402/Disulfide bonds: #status predicted  
F;223, 274, 378/Active site: His, Asp, Ser #status predicted  
F;324/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 9.9%; Score 138; DB 1; Length 433;  
Best Local Similarity 32.7%; Pred. No. 0.00054;  
Matches 32; Conservative 14; Mismatches 36; Indels 16; Gaps 4;

QY 24 CFWDNGHLYREDQTSFAPGLRCLNWLDA-----QSGLASAPVSGAGNHSCRNPDDEP 76  
DB 69 CYEGNGHFYRGKASDTMGRSCLANWSATVLOQTVAHRSALQGLGKHNYCRNPD-NR 127  
QY 77 RGPWCYVSGEAGVPEK-----RPECEDLRCPETTSQAL 108  
DB 128 RPPWCYV--QVGLKQRVQECVHNCADGKPSPEEL 163

RESULT 9  
JCS061  
macrophage-stimulating protein 1 precursor - rat  
C;Species: Rattus norvegicus (Norway rat)  
C;Date: 31-Jan-1997 #sequence\_revision 31-Jan-1997 #text\_change 16-Jun-2000  
C;Accession: JCS061  
R;Onishi, K.; Iwama, A.; Matsuno, K.; Ezaki, T.; Sakamoto, O.; Hamaguchi, I.; Takasu, N.  
Biochem. Biophys. Res. Commun. 227, 273-280, 1996  
A;Title: Molecular cloning of rat macrophage-stimulating protein and its involvement in  
A;Reference number: JCS061; MUID: 97011126; PMID: 8858136  
A;Accession: JCS061  
A;Molecule type: mRNA  
A;Residues: 1-716 <OHS>  
A;Cross-references: EMBL:X95096; NID:G1669718; PIDN:CAA64473.1; PID:G1669719  
C;Complex: disulfide-bonded heterodimer of chains derived from the same precursor  
C;Superfamily: hepatocyte growth factor; kringle homology; trypsin homology  
C;Keywords: duplication; glycoprotein; growth factor; kringle  
F;1-31/Domain: signal sequence #status predicted <SIG>  
F;32-488, 489-716/Product: macrophage-stimulating protein 1 #status predicted <MAT>  
F;32-488/Domain: macrophage-stimulating protein 1 alpha chain #status predicted <ACH>  
F;110-186/Domain: kringle homology <KRI1>  
F;191-268/Domain: kringle homology <KRI2>  
F;292-370/Domain: kringle homology <KRI3>  
F;379-457/Domain: kringle homology <KRI4>  
F;489-716/Domain: macrophage-stimulating protein 1 beta chain #status predicted <BCH>  
F;489-709/Domain: trypsin homology <TRY>  
F;72, 305, 620/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 9.9%; Score 137; DB 1; Length 716;  
Best Local Similarity 27.3%; Pred. No. 0.0012;  
Matches 48; Conservative 12; Mismatches 56; Indels 60; Gaps 8;

QY 1 LLAVQAFVSNMLLAAYGSG-----GCFWDNGHLYREDQTSFAPGLRCLNW 48  
DB 80 LLPWTO-----HSLRQLHSSLCDFQKDYVTRTIMNGASVGYRTADGLPCQAW 134  
QY 49 ---LDAQSGLASAPVSGAGNHSCRNPDDEP RGPWCYV----- 84

DB 135 SRRFPNDHKYTPFKNGL--EENFCRNPDGDPGRWCYTTNRSVRFQSGIGKSCRAVCVW 193  
QY 85 -----GEAGVPEK-RPCE--DLRCPET-----TSQALPAFTTETQASEGP 122  
DB 194 CNGEDYRGEVDVYESGREQQRWDLQHPHSHPPHPEKFPDKALKONYCRNPDASERP 249

RESULT 10  
UKPG  
u-plasminogen activator (EC 3.4.4.21.73) precursor - pig  
N;Alternate names: uPA  
C;Species: Sus scrofa domestica (domestic pig)  
C;Date: 04-Dec-1986 #sequence\_revision 17-Mar-1987 #text\_change 07-Aug-1998  
C;Accession: A00932  
R;Nagamine, Y.; Pearson, D.; Altus, M.S.; Reich, E.  
Nucleic Acids Res. 12, 9525-9541, 1984  
A;Title: cDNA and gene nucleotide sequence of porcine plasminogen activator.  
A;Reference number: A00932; MUID: 85087954; PMID: 6096832  
A;Accession: A00932  
A;Molecule type: DNA  
A;Residues: 1-240; H', 242-442 <NAG1>  
A;Experimental source: kidney cell line LLC-PK1  
R;Nagamine, Y.  
submitted to the Protein Sequence Database, December 1986  
A;Reference number: A37566  
A;Contents: annotation; correction to residue 241  
C;Genetics:  
A;Introns: 19/3; 31/1; 67/1; 125/2; 165/1; 238/2; 288/1; 335/1; 384/3  
C;Superfamily: urokinase-type plasminogen activator; EGF homology; kringle homology; tr  
C;Keywords: glycoprotein; heterodimer; hydrolase; kringle; serine proteinase  
F;1-20/Domain: signal sequence #status predicted <SIG>  
F;21-188/Product: urokinase-type plasminogen activator chain A #status predicted <ACH>  
F;33-64/Domain: EGF homology <EGF>  
F;72-153/Domain: kringle homology <KRG>  
F;150-442/Product: urokinase-type plasminogen activator chain B #status predicted <BCH>  
F;150-430/Domain: trypsin homology <TRY>  
F;152/Binding site: carbohydrate (Asn) (covalent) #status predicted  
F;179-310, 220-236, 228-299, 324-393, 356-372, 383-411/Disulfide bonds: #status predicted  
F;235, 286, 387/Active site: His, Asp, Ser #status predicted

Query Match 9.8%; Score 135.5; DB 1; Length 442;  
Best Local Similarity 36.9%; Pred. No. 0.0009;  
Matches 38; Conservative 7; Mismatches 37; Indels 21; Gaps 5;

QY 24 CFWDNGHLYREDQTSFAPGLRCLNWLDAQSGL-----ASAPVS---GAGNHSCRNPDDEP 76  
DB 72 CFEGNGHSYRGKANTNTGGRPCLPWNSATVLLNTVAHSPDALQGLGKHNYCRNPD-NQ 130  
QY 77 RGPWCYV-----GEAGVPE-----EKRPCEDLRCPETTSQ 106  
DB 131 RRPWCYVQVGLKQLVQECVHNCAGSHRPAVDGKVPFSTPE 173

RESULT 11  
JCS0599  
u-plasminogen activator (EC 3.4.4.21.68) beta precursor - common vampire bat  
N;Alternate names: tissue plasminogen activator  
C;Species: Desmodus rotundus (Common vampire bat)  
C;Date: 31-Mar-1992 #sequence\_revision 31-Mar-1992 #text\_change 16-Jul-1999  
C;Accession: JCS0599  
R;Kraetzschmar, J.; Haendler, B.; Langer, G.; Boidol, W.; Bringmann, P.; Alagon, A.; Do  
Gene 105, 229-237, 1991  
A;Title: The plasminogen activator family from the salivary gland of the vampire bat De  
A;Reference number: JCS0597; MUID: 92039036; PMID: 1937019  
A;Accession: JCS0599  
A;Molecule type: mRNA  
A;Residues: 1-431 <KEA>  
A;Cross-references: GB:M63989; NID:G166076; PIDN:AAA31594.1; PID:G166077  
C;Superfamily: tissue plasminogen activator; EGF homology; fibronectin type I repeat ho  
C;Keywords: fibrinolysis; glycoprotein; hydrolase; kringle; serine proteinase  
F;1-21/Domain: signal sequence #status predicted <SIG>  
F;22-36/Domain: propeptide #status predicted <PRO>

```

F:37-431/Product: plasminogen activator beta #status predicted <PLA>
F:41-74/Domain: EGF homology <EGF>
F:82-163/Domain: kringle homology <KR>
F:180-425/Domain: kringle homology <TRY>
F:41-52,46-63,65-74,82-163,103-145,134-158,168-299,211-227,219-288,313-388/Disulfide bond
F:139,352/Binding site: carboxylate (Asn) (covalent) #status predicted
F:179-180/Cleavage site: His-Ser (plasmin) #status predicted
F:226,275,382/Active site: His, Asp, Ser #status predicted
F:345-361,378-406/Disulfide bonds: #status predicted

Query Match          9.7%  Score 134;  DB 2;  Length 431;
Best Local Similarity 38.6%  Pred. No. 0.0012;
Matches 27;  Conservative          9;  Mismatches 20;  Indels 14;  Gaps 3;

QY  24  CFWDNGHLYREDQTSPPAGRLCNWLDAQSL-----ASAPVSGAGNSHYCRNPD 73
DB  82  CYKQDGVYTRGTWSTSSGAQCINW---NSNLLTRTYNGRRSDAITLGLGNHYCRNPD 138
QY  74  EDRPGPWCYV 83
DB  139  NNSK-PWCYV 147

RESULT 12
A34369
t-plasminogen activator (EC 3.4.21.68) precursor - false vampire bat (Megaderma lyra)
C:Species: Megaderma lyra
C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 10-Sep-1999
C:Accession: A34369
R:Gardell, S.J.; Duong, L.T.; Diehl, R.E.; York, J.D.; Hare, T.R.; Register, R.B.; Jacob
J. Biol. Chem. 264, 17947-17952, 1989
A:Title: Isolation, characterization, and cDNA cloning of a vampire bat salivary plasmin
A:Reference number: A34369; MUID:90036867; PMID:2509450
A:Accession: A34369
A:Status: Preliminary
A:Molecule type: mRNA
A:Residues: 1-477 <GAR>
A:Cross-references: GB:J05082; NID:G166080; PIDN:AAA31596.1; PID:G166081
C:Superfamily: tissue plasminogen activator; EGF homology; fibronectin type I repeat bor
C:Keywords: fibrinolysis; glycoprotein; hydrolase; kringle; serine proteinase
F:1-21/Domain: signal sequence #status predicted <SIG>
F:22-36/Domain: propeptide #status predicted <PRO>
F:37-477/Product: plasminogen activator #status predicted <PLA>
F:42-79/Domain: fibronectin type I repeat homology <1FA>
F:87-120/Domain: EGF homology <EGF>
F:128-209/Domain: kringle homology <KR>
F:226-471/Domain: trypsin homology <TRY>
F:42-72,70-79,87-98,92-109,111-120,128-209,149-191,180-204,214-345,257-273,265-334,359-4
F:272,321,428/Active site: His, Asp, Ser #status predicted

Query Match          9.7%  Score 134;  DB 1;  Length 477;
Best Local Similarity 38.6%  Pred. No. 0.0013;
Matches 27;  Conservative          9;  Mismatches 20;  Indels 14;  Gaps 3;

QY  24  CFWDNGHLYREDQTSPPAGRLCNWLDAQSL-----ASAPVSGAGNSHYCRNPD 73
DB  128  CYKQDGVYTRGTWSTSSGAQCINW---NSNLLTRTYNGRRSDAITLGLGNHYCRNPD 184
QY  74  EDRPGPWCYV 83
DB  185  NNSK-PWCYV 193

RESULT 13
JS0598
t-plasminogen activator (EC 3.4.21.68) alpha-2 precursor - common vampire bat
N:Alternate names: tissue plasminogen activator
C:Species: Desmodus rotundus (common vampire bat)
C:Date: 31-Mar-1992 #sequence_revision 31-Mar-1992 #text_change 16-Jul-1999
C:Accession: JS0598
R:Kraetzschmar, J.; Haendler, B.; Langer, G.; Boidol, W.; Bringmann, P.; Alagon, A.; Don
Gene 105, 229-237, 1991
A:Title: The plasminogen activator family from the salivary gland of the vampire bat Des

```

Db 80 LLPWTQ-----HSLHTQLYHSLCHLFQKDYVVRTCINDNGSVSYRGTVARTAGGLPCQAW 134  
QY 49 ---LDAQSGLASAPVSGAGNHSYCNPDDEPRGWCYVS----- 84  
Db 135 SRRFPNDHKYTPTPXNGL-EENFCRNPDPDPRGWCYTTRSVRFQSCGINTCREAVCVL 193  
QY 85 -----GEAGVPEK-RPCE--DLRCPET 103  
Db 194 CNGEDYRGEVDVTESGRECORWDLQPHS 222

RESULT 15  
J06000  
t-plasminogen activator (EC 3.4.21.68) gamma precursor - common vampire bat  
N:Alternate names: tissue plasminogen activator  
C:Species: Desmodus rotundus (common vampire bat)  
C:Date: 31-Mar-1992 #sequence\_revision 31-Mar-1992 #text\_change 16-Jul-1999  
C:Accession: J06000  
R:Kraetzschmar, J.; Haendler, B.; Langer, G.; Boidol, W.; Bringmann, P.; Alagon, A.; Don  
Gene 105, 229-237, 1991  
A:Title: The plasminogen activator family from the salivary gland of the vampire bat Des  
A:Reference number: J0597; MUID:92039036; PMID:1937019  
A:Accession: J06000  
A:Molecule type: mRNA  
A:Residues: 1-394 <KRA>  
A:Cross-references: GB:M63990; NID:g166078; PIDN:AAA31595.1; PID:g166079  
A:Note: The authors translated the codon ATC for residue 75 as Thr  
C:Superfamily: tissue plasminogen activator; EGF homology; fibronectin type I repeat hom  
C:Keywords: fibrinolysis; glycoprotein; hydrolase; kringle; serine proteinase  
F:1-21/Domain: signal sequence #status predicted <SIG>  
F:22-36/Domain: propeptide #status predicted <PRO>  
F:37-394/Product: plasminogen activator gamma #status predicted <PLA>  
F:45-126/Domain: kringle homology <KRG>  
F:143-388/Domain: trypsin homology <TRY>  
F:45-126,66-108,97-121,131-262,174-190,182-251,276-351,308-324,341-369/Disulfide bonds:  
F:142-143/Cleavage site: His-Ser (plasmin) #status predicted  
F:189,238,345/Active site: His, Asp, Ser #status predicted  
F:315/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 9.5%; Score 132; DB 2; Length 394;  
Best Local Similarity 32.7%; Pred. No. 0.0016;  
Matches 33; Conservative 9; Mismatches 39; Indels 20; Gaps 5;

QY 16 ABAYGS--CGCFWDNGHLVREDQTSFAPGLRCLNW-----LDAQSGLASAPVSGAGNH 66  
Db 35 SRAYGDPHATCYXQGVYIRGTWTSSESQAQINNNSNLLIRTYNGRMPFAVKLGLGNH 94

QY 67 SYCRNPDEDPGWCYV-----SGEAGVPEKRPCEDLRC 100  
Db 95 NYCRRPDGASK-PWCYVIKARKFTSBCSVP---VCSKATC 131

Search completed: March 17, 2004, 07:07:58  
Job time : 19.3365 secs

GenCore version 5.1.1.6

Copyright (C) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 17, 2004, 06:55:43 ; Search time 10.5787 Seconds  
(without alignments)  
1289.604 Million cell updates/sec

Title: US-10-057-951-2\_COPY\_2\_263

Perfect score: 1388

Sequence: 1 LLAWQAFVSNMLAEAYG.....PVDQEGSTPLMGQATPGA 262

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_42.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	160.5	11.6	566	1 TPA_BOVIN	Q28198 bos taurus
2	154.5	11.1	562	1 TPA_HUMAN	P00750 homo sapien
3	154	11.1	559	1 TPA_RAT	P19637 rattus norv
4	150	10.8	653	1 HGPA_MOUSE	Q9r098 mus musculu
5	147.5	10.6	559	1 TPA_MOUSE	P11214 mus musculu
6	146.5	10.6	655	1 HGPA_HUMAN	Q04756 homo sapien
7	142	10.2	810	1 PLMN_HUMAN	P00747 homo sapien
8	138	9.9	169	1 PLMN_RAT	Q01177 rattus norv
9	138	9.9	433	1 UROK_PIG	P16227 papio cynoc
10	135.5	9.8	442	1 UROK_PIG	P04185 sus scrofa
11	134	9.7	431	1 URTB_DESRO	P98121 desmodus ro
12	134	9.7	477	1 URTB_DESRO	P15638 desmodus ro
13	134	9.7	716	1 HGFL_MOUSE	P28928 mus musculu
14	132	9.5	394	1 URTG_DESRO	P49150 desmodus ro
15	132	9.5	431	1 UROK_HUMAN	P00749 homo sapien
16	130.5	9.4	461	1 KRW2_MOUSE	Q8K157 mus musculu
17	130	9.4	810	1 PLMN_ERIEU	Q29485 erinaceus e
18	128.5	9.3	434	1 UROK_CHICK	P15120 gallus gall
19	127	9.1	433	1 UROK_BOVIN	Q05589 bos taurus
20	126.5	9.1	728	1 HGF_MOUSE	Q03048 mus musculu
21	124.5	9.0	810	1 PLMN_MACMU	P12545 macaca mula
22	124	8.9	622	1 THRE_HUMAN	P00734 homo sapien
23	124	8.9	728	1 HGF_HUMAN	P14210 homo sapien
24	124	8.9	812	1 PLMN_BOVIN	P06868 bos taurus
25	123	8.9	593	1 FAI2_BOVIN	P98140 bos taurus
26	122.5	8.8	458	1 APOA_HUMAN	P08519 homo sapien
27	122	8.8	603	1 FAI2_CAVPO	Q04962 cavia porce
28	121.5	8.8	462	1 KRW2_HUMAN	Q8ncwo homo sapien
29	121.5	8.8	728	1 HGF_RAT	P17945 rattus norv
30	120.5	8.7	333	1 PLMN_CANFA	P80009 canis fami
31	120	8.6	433	1 UROK_MOUSE	P06869 mus musculu
32	120	8.6	790	1 PLMN_PIG	P06867 sus scrofa
33	119.5	8.6	477	1 URT1_DESRO	P98119 desmodus ro

34	118.5	8.5	615	1 FAI2_HUMAN	P00748 homo sapien
35	118	8.5	432	1 UROK_RAT	P29598 rattus norv
36	117.5	8.5	812	1 PLMN_MOUSE	P20918 mus musculu
37	117	8.4	625	1 THRE_BOVIN	P00735 bos taurus
38	116	8.4	1420	1 APOA_MACMU	P14417 macaca mula
39	115.5	8.3	711	1 HGFL_HUMAN	P26927 homo sapien
40	114	8.2	452	1 KRMI_XENLA	Q90900 xenopus lae
41	114	8.2	473	1 KRMI_MOUSE	Q99043 mus musculu
42	114	8.2	473	1 KRMI_RAT	Q92484 rattus norv
43	114	8.2	475	1 KRMI_HUMAN	Q96mu8 homo sapien
44	114	8.2	618	1 THRE_MOUSE	P19221 mus musculu
45	110.5	8.0	685	1 RORI_DROME	Q24488 drosophila

## ALIGNMENTS

RESULT 1  
TPA\_BOVIN  
ID TPA\_BOVIN STANDARD; PRT; 566 AA.  
AC Q28198;  
DT 01-NOV-1997 (Rel. 35, Created)  
DT 01-NOV-1997 (Rel. 35, Last sequence update)  
DT 15-MAR-2004 (Rel. 43, Last annotation update)  
DE Tissue-type plasminogen activator precursor (EC 3.4.21.68) (tPA)  
DE (t-PA) (t-plasminogen activator).  
GN PLAT.  
OS Bos taurus (Bovine).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;  
OC Bovidae; Bovinae; Bos.  
OX NCBI\_TaxID=9913;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Kidney;  
RA Ravn P., Berglund L., Petersen T.E.;  
RT "Cloning and characterization of the bovine plasminogen activators uPA  
and tPA";  
RL Int. Dairy J. 5:605-617(1995).  
CC -!- FUNCTION: Converts the abundant, but inactive, zymogen plasminogen  
to plasmin by hydrolyzing a single Arg-Val bond in plasminogen. By  
controlling plasmin-mediated proteolysis, it plays an important  
role in tissue remodeling and degradation, in cell migration and  
many other physiological events.  
CC -!- CATALYTIC ACTIVITY: Specific cleavage of Arg-|-Val bond in  
plasminogen to form plasmin.  
CC -!- SUBUNIT: Heterodimer of chain A and chain B held by a disulfide  
bond.  
CC -!- SUBCELLULAR LOCATION: Secreted; extracellular.  
CC -!- PTM: THE SINGLE CHAIN, ALMOST FULLY ACTIVE ENZYME, CAN BE FURTHER  
PROCESSED INTO A TWO-CHAIN FULLY ACTIVE FORM BY A CLEAVAGE AFTER  
ARG-314 CATALYZED BY PLASMIN, TISSUE KALLIKREIN OR FACTOR XA.  
CC -!- MISCELLANEOUS: Binds to the kringle structure of the fibrin A  
chain. Binding to fibrin enhances its catalytic activity.  
CC -!- SIMILARITY: Belongs to peptidase family S1.  
CC -!- SIMILARITY: Contains 1 EGF-like domain.  
CC -!- SIMILARITY: Contains 1 fibronectin type I domain.  
CC -!- SIMILARITY: Contains 2 kringle domains.

-----  
This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----

EMBL; X85900; CAA59795.1; -;  
HSSP; P00750; INTF.  
MEROPS; S01.232; -;  
InterPro; IPR009003; Cys Ser trypsin.  
InterPro; IPR006209; EGF-like.  
InterPro; IPR000083; Fibinctn1.

DR InterPro; IPR006210; IEGF.  
DR InterPro; IPR000001; Kringle.  
DR InterPro; IPR001254; Peptidase\_S1.  
DR InterPro; IPR001314; Peptidase\_S1A.  
DR Pfam; PF00008; EGF; 1.  
DR Pfam; PF00039; fnl; 1.  
DR Pfam; PF00051; kringle; 2.  
DR Pfam; PF00089; trypsin; 1.  
DR PRINTS; PR00722; CHYMOTRYPSIN.  
DR PRINTS; PR00018; KRINGLE.  
DR ProDom; PD000395; Kringle; 2.  
DR SMART; SM00181; EGF; 1.  
DR SMART; SM00058; fnl; 1.  
DR SMART; SM00130; KR; 2.  
DR SMART; SM00020; Tryp\_Spc; 1.  
DR PROSITE; PS00022; EGF\_1; 1.  
DR PROSITE; PS01186; EGF\_2; 1.  
DR PROSITE; PS00026; EGF\_3; 1.  
DR PROSITE; PS01253; FIBRONECTIN\_1; 1.  
DR PROSITE; PS00021; KRINGLE\_1; 1.  
DR PROSITE; PS00070; KRINGLE\_2; 1.  
DR PROSITE; PS00240; TRYPSIN\_DOM; 1.  
DR PROSITE; PS00134; TRYPSIN\_HIS; 1.  
DR PROSITE; PS00135; TRYPSIN\_SER; 1.  
KW Plasminogen activator; Hydrolase; Serine protease; Glycoprotein;  
KW Plasma; Kringle; EGF-like domain; Repeat; Signal.  
FT SIGNAL 1 21  
FT PROPRP 22 33 BY SIMILARITY.  
FT CHAIN 34 566 TISSUE-TYPE PLASMINOGEN ACTIVATOR.  
FT CHAIN 34 314 TISSUE-TYPE PLASMINOGEN ACTIVATOR A  
FT CHAIN 315 566 TISSUE-TYPE PLASMINOGEN ACTIVATOR B  
FT DOMAIN 40 82 FIBRONECTIN TYPE-1.  
FT DOMAIN 83 121 EGF-LIKE.  
FT DOMAIN 128 209 KRINGLE 1.  
FT DOMAIN 219 300 KRINGLE 2.  
FT DOMAIN 315 566 SERINE PROTEASE.  
FT ACT\_SITE 361 361 CHARGE RELAY SYSTEM.  
FT ACT\_SITE 410 410 CHARGE RELAY SYSTEM.  
FT ACT\_SITE 517 517 CHARGE RELAY SYSTEM.  
FT DISULFID 42 72 BY SIMILARITY.  
FT DISULFID 70 79 BY SIMILARITY.  
FT DISULFID 87 98 BY SIMILARITY.  
FT DISULFID 92 109 BY SIMILARITY.  
FT DISULFID 111 120 BY SIMILARITY.  
FT DISULFID 128 209 BY SIMILARITY.  
FT DISULFID 149 191 BY SIMILARITY.  
FT DISULFID 180 204 BY SIMILARITY.  
FT DISULFID 219 300 BY SIMILARITY.  
FT DISULFID 240 282 BY SIMILARITY.  
FT DISULFID 271 295 BY SIMILARITY.  
FT DISULFID 303 434 INTERCHAIN (BY SIMILARITY).  
FT DISULFID 346 362 BY SIMILARITY.  
FT DISULFID 354 423 BY SIMILARITY.  
FT DISULFID 448 523 BY SIMILARITY.  
FT DISULFID 480 496 BY SIMILARITY.  
FT DISULFID 513 541 BY SIMILARITY.  
FT CARBOHYD 153 153 N-LINKED (GLCNAC... ) (POTENTIAL).  
FT CARBOHYD 487 487 N-LINKED (GLCNAC... ) (POTENTIAL).  
SQ SEQUENCE 566 AA; 63701 MW; 2868BEB4E32276C3 CRC64;  
Query Match 11.6%; Score 160.5; DB 1; Length 566;  
Best Local Similarity 35.5%; Pred. No. 5.6e-06;  
Matches 38; Conservative 11; Mismatches 45; Indels 13; Gaps 4;  
Qy 15 LAEAYGGGFWNGHLYRDTQSPAPGLCLNWLDAQSLAPVY-----GAGN 65  
Db 119 LCEIDATATCYQDQGVAYRGVTWSTAESGAECANW--NSSGLAMKPYSGRRPNAILGLGN 176  
Qy 66 HSYCRNDEPRGFWCVSGEAGVPEKRPCEDLRCPTTSCALPAFT 112  
Db 177 HNYCRNPDQSK-FWCIVF-KAGYISEPCSTACAKVAEDGDCYT 221

RESULT 2  
TPA\_HUMAN STANDARD; PRT; 562 AA.  
AC P00750; Q15103;  
DT 21-JUL-1986 (Rel. 01, Created)  
DT 21-JUL-1986 (Rel. 01, Last sequence update)  
DT 15-MAR-2004 (Rel. 43, Last annotation update)  
DE Tissue-type plasminogen activator precursor (EC 3.4.21.68) (TPA)  
DE (t-PA) (t-plasminogen activator) (Alteplase) (Rateplase).  
GN PLAT.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.  
OC NCBI\_TaxID=9606;  
RX [1]  
SEQUENCE FROM N.A.  
RC TISSUE-Melanoma;  
RX MEDLINE=83115262; PubMed=6337343;  
RA Pennica D., Holmes W.E., Kohr W.J., Harkins R.N., Vehar G.A.,  
Ward C.A., Bennett W.F., Yelverton E., Seeburg P.H., Heynesker H.L.,  
Goeddel D.V., Collen D.;  
RA "Cloning and expression of human tissue-type plasminogen activator  
RT cDNA in E. coli.";  
RL Nature 301:214-221(1983).  
[2]  
SEQUENCE FROM N.A.  
RC TISSUE-Fetal lung;  
RX MEDLINE=88262579; PubMed=3133640;  
RA Sasaki H., Saito Y., Hayashi M., Otsuka K., Niwa M.;  
RT "Nucleotide sequence of the tissue-type plasminogen activator cDNA  
RL from human fetal lung cells.";  
RL Nucleic Acids Res. 16:5695-5695(1988).  
[3]  
SEQUENCE FROM N.A.  
RX MEDLINE=88054470; PubMed=2824147;  
RA Reddy V.B., Garrazone A.J., Sasak H., Wei C.-M., Watkins P., Galli J.,  
Hsiung N.;  
RT "Expression of human uterine tissue-type plasminogen activator in  
RL mouse cells using BPV vectors.";  
RL DNA 6:461-472(1987).  
[4]  
SEQUENCE FROM N.A.  
RX MEDLINE=86196143; PubMed=3009482;  
RA Frieznher Degen S.J., Rajput B., Reich E.;  
RT "The human tissue plasminogen activator gene.";  
RL J. Biol. Chem. 261:6972-6985(1986).  
[5]  
SEQUENCE FROM N.A.  
RX MEDLINE=84298137; PubMed=6089198;  
RA Ny T., Elgh F., Lund B.;  
RT "The structure of the human tissue-type plasminogen activator gene:  
RT correlation of intron and exon structures to functional and  
RT structural domains.";  
RL Proc. Natl. Acad. Sci. U.S.A. 81:5355-5359(1984).  
[6]  
SEQUENCE FROM N.A.  
RX MEDLINE=86284200; PubMed=3090401;  
RA Harris T.J., Patel I., Marston F.A., Little S., Emtage J.S.,  
RA Opdenakker G., Volckaert G., Rombauts W., Billiau A., Somer P.;  
RT "Cloning of cDNA coding for human tissue-type plasminogen activator  
RT and its expression in Escherichia coli.";  
RL Mol. Biol. Med. 3:279-292(1986).  
[7]  
SEQUENCE FROM N.A. (ISOFORM SHORT).  
RC TISSUE-Umbilical vein;  
RX MEDLINE=90192129; PubMed=2107528;  
RA Siebert P.D., Fong K.;  
RT "Variant tissue-type plasminogen activator (PLAT) cDNA obtained from  
RL human endothelial cells.";  
RL Nucleic Acids Res. 18:1086-1086(1990).  
[8]

RP SEQUENCE FROM N.A.  
 RC TISSUE=Brain;  
 RX MEDLINE=2238257; PubMed=12477932;  
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
 RA Klautner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,  
 RA Altshuler S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,  
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,  
 RA Diatchenko L., Marusik K., Farmer A.A., Rubin G.M., Hong L.,  
 RA Stapleton M.J., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.B.,  
 RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,  
 RA Raha S.S., Loeuillano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,  
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Rulyk S.W.,  
 RA Villalón D.K., Muzny R.C., Sodergren E.J., Lu X., Gibbs R.A.,  
 RA Fahy J., Helton E., Kettner M., Madan A., Rodriguez S., Sanchez A.,  
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,  
 RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,  
 RA Schnerth A., Schein J.E., Jones S.J.M., Marra M.A.;  
 RT "Generation and initial analysis of more than 15,000 full-length  
 RT human and mouse cDNA sequences.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).  
 RN [9]  
 RP SEQUENCE OF 212-361 FROM N.A.  
 RX MEDLINE=83169656; PubMed=6572897;  
 RA Edlund T., Ny T., Raanby M., Heden L.-O., Palm G., Holmgren E.,  
 RA Josephson S.;  
 RT "Isolation of cDNA sequences coding for a part of human tissue  
 RT plasminogen activator.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 80:349-352(1983).  
 RN [10]  
 RP SEQUENCE OF 1-36 FROM N.A.  
 RX MEDLINE=85289338; PubMed=3161893;  
 RA Fisher R., Waller E.K., Grossi G., Thompson D., Tizard R.,  
 RA Schleuning W.-D.;  
 RT "Isolation and characterization of the human tissue-type plasminogen  
 RT activator structural gene including its 5' flanking region.";  
 RL J. Biol. Chem. 260:11223-11230(1985).  
 RN [11]  
 RP SEQUENCE OF 31-562 FROM N.A.  
 RX MEDLINE=91291340; PubMed=1368681;  
 RA Itagaki Y., Yasuda H., Morinaga T., Mitsuda S., Higashio K.;  
 RA "Purification and characterization of tissue plasminogen activator  
 RT secreted by human embryonic lung diploid fibroblasts, IMR-90 cells.";  
 RL Agric. Biol. Chem. 55:1225-1232(1991).  
 RN [12]  
 RP SEQUENCE OF 36-562.  
 RC TISSUE=Melanoma;  
 RX MEDLINE=85000468; PubMed=6433976;  
 RA Pohl G., Kaellstroem M., Bergsdorf N., Wallen P., Joernvall H.;  
 RA "Tissue plasminogen activator: peptide analyses confirm an indirectly  
 RT derived amino acid sequence, identify the active site serine residue,  
 RT establish glycosylation sites, and localize variant differences.";  
 RL Biochemistry 23:3701-3707(1984).  
 RN [13]  
 RP SEQUENCE OF 33-52 AND 311-330.  
 RC TISSUE=Melanoma;  
 RX MEDLINE=83209620; PubMed=6682760;  
 RA Wallen P., Pohl G., Bergsdorf N., Raanby M., Ny T., Joernvall H.;  
 RA "Purification and characterization of a melanoma cell plasminogen  
 RT activator.";  
 RL Eur. J. Biochem. 132:681-686(1983).  
 RN [14]  
 RP STRUCTURE OF CARBOHYDRATES.  
 RX MEDLINE=90092112; PubMed=2513186;  
 RA Pfeiffer G., Schmidt M., Strube K.-H., Geyer R.;  
 RT "Carbohydrate structure of recombinant human urokinase tissue  
 RT plasminogen activator expressed in mouse epithelial cells.";  
 RL Eur. J. Biochem. 186:273-286(1989).  
 RN [15]  
 RP CARBOHYDRATE-LINKAGE SITE THR-96.  
 RX MEDLINE=91159408; PubMed=1900431;

RA Harris R.J., Leonard C.K., Guzzetta A.W., Spellman M.W.;  
 RT "Tissue plasminogen activator has an O-linked fucose attached to  
 RT threonine-61 in the epidermal growth factor domain.";  
 RL Biochemistry 30:2311-2314(1991).  
 RN [16]  
 RP DISULFIDE BONDS IN KRINGLE 2.  
 RX MEDLINE=91244765; PubMed=1645336;  
 RA Vlahos C.J., Wilhelm O.G., Hassell T., Jaskunas S.R., Bang N.U.;  
 RA "Disulfide pairing of the recombinant kringle-2 domain of tissue  
 RT plasminogen activator produced in *Escherichia coli*.";  
 RL J. Biol. Chem. 266:10070-10072(1991).  
 RN [17]  
 RP X-RAY CRYSTALLOGRAPHY (2.3 ANGSTROMS) OF CATALYTIC DOMAIN.  
 RX MEDLINE=96200985; PubMed=8613982;  
 RA Lamba D., Bauer M., Huber R., Fischer S., Rudolph R., Kohnert U.,  
 RA Bode W.;  
 RT "The 2.3 Å crystal structure of the catalytic domain of recombinant  
 RT two-chain human tissue-type plasminogen activator.";  
 RL J. Mol. Biol. 258:117-135(1996).  
 RN [18]  
 RP X-RAY CRYSTALLOGRAPHY (3.1 ANGSTROMS) OF CATALYTIC DOMAIN.  
 RX MEDLINE=97449126; PubMed=9305622;  
 RA Renatus M., Engh R.A., Stubbs M.T., Huber R., Fischer S., Kohnert U.,  
 RA Bode W.;  
 RT "Lysine 156 promotes the anomalous proenzyme activity of tPA: X-ray  
 RT crystal structure of single-chain human tPA.";  
 RL EMBO J. 16:4797-4805(1997).  
 RN [19]  
 RP X-RAY CRYSTALLOGRAPHY (2.4 ANGSTROMS) OF KRINGLE 2.  
 RX MEDLINE=92118803; PubMed=1310033;  
 RA de Vos A., Ullsch M.H., Kelley R.F., Padmanabhan K., Tulinsky A.,  
 RA Westbrook M.L., Kossiakof A.A.;  
 RT "Crystal structure of the kringle 2 domain of tissue plasminogen  
 RT activator at 2.4-Å resolution.";  
 RL Biochemistry 31:270-279(1992).  
 RN [20]  
 RP STRUCTURE BY NMR OF KRINGLE 2.  
 RX MEDLINE=90122799; PubMed=2558718;  
 RA Byeon I.-J.L., Kelley R.F., Llinas M.;  
 RT "1H NMR structural characterization of a recombinant kringle 2 domain  
 RT from human tissue-type plasminogen activator.";  
 RL Biochemistry 28:9350-9360(1989).  
 RN [21]  
 RP STRUCTURE BY NMR OF KRINGLE 2.  
 RX MEDLINE=91200042; PubMed=1901789;  
 RA Byeon I.-J.L., Kelley R.F., Llinas M.;  
 RT "Kringle-2 domain of the tissue-type plasminogen activator. 1H-NMR  
 RT assignments and secondary structure.";  
 RL Eur. J. Biochem. 197:155-165(1991).  
 RN [22]  
 RP STRUCTURE BY NMR OF KRINGLE 2.  
 RX MEDLINE=92106329; PubMed=1762144;  
 RA Byeon I.-J.L., Llinas M.;  
 RT "Solution structure of the tissue-type plasminogen activator kringle  
 RT 2 domain complexed to 6-aminohexanoic acid an antifibrinolytic  
 RT drug.";  
 RL J. Mol. Biol. 222:1035-1051(1991).  
 RN [23]

Query Match 11.1%; Score 154.5; DB 1; Length 562;  
 Best Local Similarity 39.6%; Pred. No. 1.8e-05;  
 Matches 36; Conservative 6; Mismatches 36; Indels 13; Gaps 4;  
 Qy 24 CFWDNGHLYREDQTSFAPGLRCLNWLDAOSGLASAPVS-----GAGNHSYCRNPDE 74  
 Db 127 CYEDQGISYRGTWSTAESGAECTNW--NSSALAKPYSGRRPDAIRLGLGNHYCRNPDR 184  
 Qy 75 DPRGPACYSGEAGVPEKPCEDLRCPETTS 105  
 Db 185 DSK-PWCYVF-KAGKYSSEFCSTPACSEGS 213  
 RESULT 3

DR	MEROPS; S01_232;
DR	InterPro; IPR003003; Cys_Ser_trypsin.
DR	InterPro; IPR006209; EGF_like.
DR	InterPro; IPR000083; Fibrnctnl.
DR	InterPro; IPR006210; HEGF.
DR	InterPro; IPR000001; Kringle.
DR	InterPro; IPR001254; Peptidase_S1.
DR	InterPro; IPR001314; Peptidase_S1A.
DR	Pfam; PF00008; EGF; 1.
DR	Pfam; PF00039; fnl; 1.
DR	Pfam; PF00051; kringle; 2.
DR	Pfam; PF00089; trypsin; 1.
DR	PRINTS; PR00722; CHYMOTRYPSIN.
DR	PRINTS; PR00018; KRINGLE.
DR	ProDom; PD000395; Kringle; 2.
DR	SMART; SM00181; EGF; 1.
DR	SMART; SM00058; kringle; 1.
DR	SMART; SM00130; KR; 2.
DR	SMART; SM00020; Tryp_SPC; 1.
DR	PROSITE; PS00022; EGF-1; 1.
DR	PROSITE; PS01186; EGF-2; 1.
DR	PROSITE; PS00026; EGF-3; 1.
DR	PROSITE; PS01253; FIBRONECTIN_1; 1.
DR	PROSITE; PS00021; KRINGLE_1; 2.
DR	PROSITE; PS00070; KRINGLE_2; 2.
DR	PROSITE; PS02040; TRYPSIN_DOM; 1.
DR	PROSITE; PS00134; TRYPSIN_HIS; 1.
DR	PROSITE; PS00135; TRYPSIN_SER; 1.
DR	KW Plasminogen activation; Hydrolase; Serine protease; Glycoprotein; plasma; Kringle; EGF-like domain; Repeat; Signal. PROBABLE.
FT	SIGNAL 1 17
FT	PROPSP 18 29
FT	CHAIN 30 559
FT	CHAIN 30 308
FT	CHAIN 309 559
FT	DOMAIN 36 78
FT	DOMAIN 79 117
FT	DOMAIN 124 205
FT	DOMAIN 213 294
FT	DOMAIN 309 559
FT	ACT_SITE 355 355
PT	ACT_SITE 404 404
PT	ACT_SITE 510 510
FT	DISULFID 38 68
FT	DISULFID 66 75
FT	DISULFID 83 94
FT	DISULFID 88 105
FT	DISULFID 107 116
FT	DISULFID 124 205
FT	DISULFID 145 187
FT	DISULFID 176 200
FT	DISULFID 213 294
FT	DISULFID 234 276
FT	DISULFID 265 289
FT	DISULFID 297 428
FT	DISULFID 340 356
FT	DISULFID 348 417
FT	DISULFID 442 516
FT	DISULFID 474 490
FT	DISULFID 506 534
FT	CARBOHYD 149 149
FT	CARBOHYD 481 481
FT	CONFLICT 380 380
FT	SEQUENCE 559 AA; 62903 MW; 7DBD3809C1D1921 CRC64;
FT	Query Match 11.1%; Score 154; DB 1; Length 559;
FT	Best Local Similarity 32.4%; Pred No. 1.ee-05;
FT	Matches 45; Conservative 14; Mismatches 54; Indels 26;



Db 124 CPEGQITVGTWSTANGAECINW--NSSALSQKPYSPARRNAIKLGNHNYCRNDR 181

QY 75 DRPGPCWYSGVAGVPEKPCEDLRCP-----TTSQALPAFTTEIQEASGPG 123

Db 182 DVK-PWCYTF-KAGKYTFEFCSTPACPKGFTDCYVKGVTYRTHSFTT--SKASCLPW 237

QY 124 ADEVQVAFANALPARESEA 142

Db 238 NSMILIGKTYTAWANSQA 256

RESULT 4

HGFA MOUSE

ID HGFA MOUSE STANDARD; PRT; 653 AA.

AC Q9R038: Q9JKV4;

DT 16-OCT-2001 (Rel. 40, Created)

DT 16-OCT-2001 (Rel. 40, Last sequence update)

DT 15-MAR-2004 (Rel. 43, Last annotation update)

DE Hepatocyte growth factor activator precursor (EC 3.4.21.-) (HGF activator) (HGFA).

GN HGFA.

OS Mus musculus (Mouse).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

OX NCBI\_TaxID=10090;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=BALB/c;

RA Itoh H., Kataoka H., Koono H.;

RT "Mouse hepatocyte growth factor activator.";

RL Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.

[2]

RP SEQUENCE FROM N.A.

RX MEDLINE=21226753; PubMed=11032833;

RA van Adelsberg J.S., Sehgal S., Kukes A., Brady C., Barasch J., Yang J., Huan Y.;

RT "Activation of hepatocyte growth factor (HGF) by endogenous HGF activator is required for metaphoric kidney morphogenesis in vitro.";

RL J. Biol. Chem. 276:15099-15106(2001).

CC -!- FUNCTION: Activates hepatocyte growth factor (HGF) by converting it from a single chain to a heterodimeric form (By similarity).

CC -!- SUBUNIT: Dimer of a short chain and a long chain linked by a disulfide bond (By similarity).

CC -!- SUBCELLULAR LOCATION: Secreted as an inactive single-chain precursor and is then activated to a heterodimeric form (By similarity).

CC -!- SIMILARITY: Belongs to peptidase family S1.

CC -!- SIMILARITY: Contains 2 EGF-like domains.

CC -!- SIMILARITY: Contains 1 fibronectin type I domain.

CC -!- SIMILARITY: Contains 1 fibronectin type II domain.

CC -!- SIMILARITY: Contains 1 kringle domain.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----

DR EMBL; AF099017; AAF02489.1; -

DR EMBL; AF224724; AAF34712.1; -

DR HSP; P00763; IDPO.

DR MEROPS; S01.228; -

DR MGD; MGI:1859281; Hgfac.

DR InterPro; IPR005003; Cys\_ser\_trypsin.

DR InterPro; IPR000742; EGF 2.

DR InterPro; IPR006209; EGF like.

DR InterPro; IPR000083; Fibnctnl.

DR InterPro; IPR000562; FN Type-II.

DR InterPro; IPR006210; IEGF.

DR InterPro; IPR000001; Kringle.

DR InterPro; IPR001254; Peptidase S1.

DR InterPro; IPR001314; Peptidase\_S1a.

DR Pfam; PF00008; EGF; 2.

DR Pfam; PF00039; fn1; 1.

DR Pfam; PF00040; fn2; 1.

DR Pfam; PF00051; kringle; 1.

DR Pfam; PF00089; trypsin; 1.

DR PRINTS; PR00722; CHYMOTRYPSIN.

DR PRINTS; PR00013; FNTYPEII.

DR PRINTS; PR00018; KRINGLE.

DR ProDom; PD000995; FN\_Type\_II; 1.

DR ProDom; PD000395; Kringle; 1.

DR SMART; SM00181; EGF; 2.

DR SMART; SM00059; FN2; 1.

DR SMART; SM00130; KR; 1.

DR SMART; SM00020; Tryp\_Spc; 1.

DR PROSITE; PS00022; EGF\_1; 2.

DR PROSITE; PS01186; EGF\_2; 1.

DR PROSITE; PS00026; EGF\_3; 2.

DR PROSITE; PS01253; FIBRONECTIN\_1; 1.

DR PROSITE; PS00023; FIBRONECTIN\_2; 1.

DR PROSITE; PS00021; KRINGLE\_1; 1.

DR PROSITE; PS00070; KRINGLE\_2; 1.

DR PROSITE; PS02040; TRYPSIN\_DOM; 1.

DR PROSITE; PS00134; TRYPSIN\_HIS; 1.

DR PROSITE; PS00135; TRYPSIN\_SER; 1.

KW Hydrolase; Glycoprotein; Plasma; Serine protease; Kringle; Signal; EGF-like domain; Repeat; Zymogen.

FT SIGNAL 1 29 BY SIMILARITY.

FT PROPEP 30 369 CLEAVED IN ACTIVE FORM (BY SIMILARITY).

FT CHAIN 370 405 HEPATOCYTE GROWTH FACTOR ACTIVATOR SHORT CHAIN.

FT CHAIN 406 653 HEPATOCYTE GROWTH FACTOR ACTIVATOR LONG CHAIN.

FT DOMAIN 105 145 FIBRONECTIN TYPE-II.

FT DOMAIN 157 195 EGF-LIKE 1.

FT DOMAIN 197 237 FIBRONECTIN TYPE-I.

FT DOMAIN 238 276 EGF-LIKE 2.

FT DOMAIN 283 364 KRINGLE.

FT DOMAIN 406 653 SERINE PROTEASE.

FT ACT\_SITE 445 445 CHARGE RELAY SYSTEM (BY SIMILARITY).

FT ACT\_SITE 495 495 CHARGE RELAY SYSTEM (BY SIMILARITY).

FT ACT\_SITE 596 596 CHARGE RELAY SYSTEM (BY SIMILARITY).

FT DISULFID 105 130 BY SIMILARITY.

FT DISULFID 119 145 BY SIMILARITY.

FT DISULFID 161 172 BY SIMILARITY.

FT DISULFID 166 183 BY SIMILARITY.

FT DISULFID 185 194 BY SIMILARITY.

FT DISULFID 199 227 BY SIMILARITY.

FT DISULFID 225 234 BY SIMILARITY.

FT DISULFID 242 253 BY SIMILARITY.

FT DISULFID 247 264 BY SIMILARITY.

FT DISULFID 266 275 BY SIMILARITY.

FT DISULFID 283 364 BY SIMILARITY.

FT DISULFID 304 346 BY SIMILARITY.

FT DISULFID 335 359 BY SIMILARITY.

FT DISULFID 392 519 INTERCHAIN (BY SIMILARITY).

FT DISULFID 430 446 BY SIMILARITY.

FT DISULFID 438 508 BY SIMILARITY.

FT DISULFID 533 602 BY SIMILARITY.

FT DISULFID 565 581 BY SIMILARITY.

FT DISULFID 592 620 BY SIMILARITY.

FT CARBOHYD 39 39 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 47 47 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 63 63 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 287 287 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 466 466 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 544 544 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CONFLICT 164 164 G -> W (IN REF. 2).

SQ SEQUENCE 653 AA; 70567 MW; 88B4B2025DF7DC CRC64;

Query Match 10.8%; Score 150; DB 1; Length 653;

Best Local Similarity 32.6%; Pred. No. 5e-05;

Matches	46;	Conservative	17;	Mismatches	46;	Indels	32;	Gaps	7;
QY	24	CFWDNGHLYREDQTPAPGLRCINW-----LDAQSLASAPVSGAGNHSYCRNPDPD	75						
DB	283	CFLNGTEYRGVASTAASGLSLAWNSDLLYQELHVDLS-VAAVLLGLGHPAYCRNPDKD	341						
QY	76	PRGPWCYVSGEAGVPEKPCEDLRCPETTSQALPAFTTEIOBASEGPGADEVQVAFAPANA	135						
DB	342	ER-PWCYVVKDNLNWE-----YCLTACSLARVHSQTPE-----ILA---A	380						
QY	136	LPARSEAAAVQPVIGISQVR	156						
DB	381	LP--ESAPAVRPTCGKRHKR	399						
RESULT 5									
TPA_MOUSE									
ID	TPA_MOUSE	STANDARD;	PRT;	559	AA.				
AC	P11214; Q91VP2;								
DT	01-JUL-1989 (Rel. 11, Created)								
DT	10-OCT-2003 (Rel. 42, Last sequence update)								
DT	15-MAR-2004 (Rel. 43, Last annotation update)								
DE	Tissue-type plasminogen activator precursor (EC 3.4.21.68) (tPA)								
DE	(t-PA) (t-plasminogen activator).								
GN	Plat.								
OS	Mus musculus (Mouse).								
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;								
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.								
OX	NCBI_TaxID=10090;								
RN	[1]								
RP	SEQUENCE FROM N.A.								
RX	MEDLINE=88087303; PubMed=2826484;								
RA	Rickles R.J., Darrow A.L., Strickland S.;								
RT	"Molecular cloning of complementary DNA to mouse tissue plasminogen								
RT	activator mRNA and its expression during F9 teratocarcinoma cell								
RT	differentiation."								
RL	J. Biol. Chem. 263:1563-1569(1988).								
RN	[2]								
RP	SEQUENCE FROM N.A.								
RC	TISSUE=Mammary gland;								
RX	MEDLINE=22388257; PubMed=12477932;								
RA	Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,								
RA	Klausner R.D., Collins F.S., Wagner K.H., Shenmen C.M., Schuler G.D.,								
RA	Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,								
RA	Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,								
RA	Datchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,								
RA	Scapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,								
RA	Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,								
RA	Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,								
RA	Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,								
RA	Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,								
RA	Vallalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,								
RA	Fahey J., Helton E., Kettelman M., Madan A., Rodrigues S., Sanchez A.,								
RA	Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,								
RA	Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,								
RA	Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,								
RA	Butterfield F.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,								
RA	Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;								
RT	"Generation and initial analysis of more than 15,000 full-length human								
RT	and mouse cDNA sequences."								
RL	Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).								
CC	-!- FUNCTION: Converts the abundant, but inactive, zymogen plasminogen								
CC	to plasmin by hydrolyzing a single Arg-Val bond in plasminogen. By								
CC	controlling plasmin-mediated proteolysis, it plays an important								
CC	role in tissue remodeling and degradation, in cell migration and								
CC	many other physiological events.								
CC	-!- CATALYTIC ACTIVITY: Specific cleavage of Arg- -Val bond in								
CC	plasminogen to form plasmin.								
CC	-!- SUBUNIT: Heterodimer of chain A and chain B held by a disulfide								
CC	bond.								
CC	-!- SUBCELLULAR LOCATION: Secreted; extracellular.								
CC	-!- PTM: THE SINGLE CHAIN, ALMOST FULLY ACTIVE ENZYME, CAN BE FURTHER								
CC	PROCESSED INTO A TWO-CHAIN FULLY ACTIVE FORM BY A CLEAVAGE AFTER								

CC	ARG-308 CATALYZED BY PLASMIN. TISSUE KALLIKREIN OR FACTOR XA.
CC	-!- MISCELLANEOUS: Binds to the kringle structure of the fibrin A
CC	chain. Binding to fibrin enhances its catalytic activity.
CC	-!- SIMILARITY: Belongs to peptidase family S1.
CC	-!- SIMILARITY: Contains 1 EGF-like domain.
CC	-!- SIMILARITY: Contains 1 fibronectin type I domain.
CC	-!- SIMILARITY: Contains 2 kringle domains.
CC	-----
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation
CC	the European Bioinformatics Institute. There are no restrictions on its
CC	use by non-profit institutions as long as its content is in no way
CC	modified and this statement is not removed. Usage by and for commercial
CC	entities requires a license agreement (See <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a>
CC	or send an email to <a href="mailto:license@sib-sib.ch">license@sib-sib.ch</a> ).
CC	-----
CC	EMBL; J03520; AAA40470.1; --
CC	EMBL; BC011256; AAH11256.1; --
CC	PIR; A29941; A29941.
CC	HSP; P00750; IASH.
CC	MEROPS; S01.232; --
CC	MGD; MGI:97610; Plat.
CC	InterPro: IPR009003; Cys_Ser_trypsin.
CC	InterPro: IPR006209; EGF-like.
CC	InterPro: IPR000083; Fibrinctn1.
CC	InterPro: IPR006210; IEGF.
CC	InterPro: IPR000001; Kringle.
CC	InterPro: IPR001254; Peptidase S1.
CC	InterPro: IPR001314; Peptidase_S1A.
CC	Pfam; PF00008; EGF; 1.
CC	Pfam; PF00039; fn1; 1.
CC	Pfam; PF00051; kringle; 2.
CC	Pfam; PF00089; trypsin; 1.
CC	PRINTS; PRO0722; CHYMOTRYPSIN.
CC	PRINTS; PRO0018; KRINGLE.
CC	ProDom; PD000395; Kringle; 2.
CC	SMART; SM00181; EGF; 1.
CC	SMART; SM00058; FN1; 1.
CC	SMART; SM00130; KR; 2.
CC	SMART; SM00020; TRY_SPC; 1.
CC	PROSITE; PS00022; EGF_1; 1.
CC	PROSITE; PS01186; EGF_2; 1.
CC	PROSITE; PS00026; EGF_3; 1.
CC	PROSITE; PS01253; FIBRONECTIN_1; 1.
CC	PROSITE; PS00021; KRINGLE_1; 2.
CC	PROSITE; PS00070; KRINGLE_2; 2.
CC	PROSITE; PS00240; TRYPSIN_DOM; 1.
CC	PROSITE; PS00134; TRYPSIN_HIS; 1.
CC	PROSITE; PS00135; TRYPSIN_SER; 1.
CC	Plasma; Kringle; EGF-like domain; Serine protease; Glycoprotein;
CC	Plasminogen activation; Hydrolase; Repeat; Signal.
CC	SIGNAL 1 17 PROBABLE.
CC	PROPEP 18 29
CC	CHAIN 30 559
CC	CHAIN 30 308
CC	-----
CC	CHAIN 309 559
CC	-----
CC	DOMAIN 36 78
CC	DOMAIN 79 117
CC	DOMAIN 124 205
CC	DOMAIN 213 294
CC	DOMAIN 309 559
CC	ACT_SITE 355 355
CC	ACT_SITE 404 404
CC	ACT_SITE 510 510
CC	DISULFID 38 68
CC	DISULFID 66 75
CC	DISULFID 83 94
CC	DISULFID 88 105
CC	DISULFID 107 116
CC	DISULFID 124 205
CC	DISULFID 145 187

CC	ARG-308 CATALYZED BY PLASMIN, TISSUE KALLIKREIN OR FACTOR XA.								
CC	-!- MISCELLANEOUS: Binds to the kringle structure of the fibrin A								
CC	chain. Binding to fibrin enhances its catalytic activity.								
CC	-!- SIMILARITY: Belongs to peptidase family S1.								
CC	-!- SIMILARITY: Contains 1 EGF-like domain.								
CC	-!- SIMILARITY: Contains 1 fibronectin type I domain.								
CC	-!- SIMILARITY: Contains 2 kringle domains.								
CC	-----								
CC	THIS SWISS-PROT entry is copyright. It is produced through a collaboration								
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -								
CC	the European Bioinformatics Institute. There are no restrictions on its								
CC	use by non-profit institutions as long as its content is in no way								
CC	modified and this statement is not removed. Usage by and for commercial								
CC	entities requires a license agreement (See <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a>								
CC	or send an email to <a href="mailto:license@isb-sib.ch">license@isb-sib.ch</a> ).								
CC	-----								
DR	EMBL; J03520; AAA04070.1; -.								
DR	EMBL; BC011256; AAH11256.1; -.								
DR	PIR; A29941; A29941.								
DR	HSP; P00750; IA5H.								
DR	MEROPS; S01.232; -.								
DR	MGD; MGI:97610; Plat.								
DR	InterPro; IPR009003; Cys Ser trypsin.								
DR	InterPro; IPR006209; EGF-like.								
DR	InterPro; IPR000083; Fibnctn1.								
DR	InterPro; IPR006210; IEGF.								
DR	InterPro; IPR000001; Kringle.								
DR	InterPro; IPR001254; Peptidase_S1.								
DR	InterPro; IPR001314; Peptidase_S1A.								
DR	Pfam; PF00008; EGF; 1.								
DR	Pfam; PF00039; fn1; 1.								
DR	Pfam; PF00051; kringle; 2.								
DR	Pfam; PF00089; trypsin; 1.								
DR	PRINTS; PR00722; CHYMOTRYPSIN.								
DR	PRINTS; PR00018; KRINGLE.								
DR	PRODOM; PD000395; Kringle; 2.								
DR	SMART; SM00181; EGF; 1.								
DR	SMART; SM00058; FN1; 1.								
DR	SMART; SM00130; KR; 2.								
DR	SMART; SMC0020; TRYP_SPC; 1.								
DR	PROSITE; PS00022; EGF_1; 1.								
DR	PROSITE; PS01186; EGF_2; 1.								
DR	PROSITE; PS00026; EGF_3; 1.								
DR	PROSITE; PS01253; FIBRONECTIN_1; 1.								
DR	PROSITE; PS00021; KRINGLE_1; 2.								
DR	PROSITE; PS00070; KRINGLE_2; 2.								
DR	PROSITE; PS0240; TRYP_SIN_DOM; 1.								
DR	PROSITE; PS00134; TRYP_SIN_HIS; 1.								
DR	PROSITE; PS00135; TRYP_SIN_SER; 1.								
KW	Plasminogen activation; Hydrolase; Serine protease; Glycoprotein;								
KW	Plasma; Kringle; EGF-like domain; Repeat; Signal.								
FT	SIGNAL	1	17						
FT	PROPEP	18	29						
FT	CHAIN	30	559						
FT	CHAIN	30	308						
FT	CHAIN	309	559						
FT	DOMAIN	36	78						
FT	DOMAIN	79	117						
FT	DOMAIN	124	205						
FT	DOMAIN	213	294						
FT	DOMAIN	309	359						
FT	ACT_SITE	355	355						
FT	ACT_SITE	404	404						
FT	ACT_SITE	510	510						
FT	DISULFID	38	68						
FT	DISULFID	66	75						
FT	DISULFID	83	94						
FT	DISULFID	88	105						
FT	DISULFID	107	116						
FT	DISULFID	124	205						
FT	DISULFID	145	187						

FT DISULFID 176 200 BY SIMILARITY.  
FT DISULFID 213 294 BY SIMILARITY.  
FT DISULFID 234 276 BY SIMILARITY.  
FT DISULFID 265 289 BY SIMILARITY.  
FT DISULFID 297 428 INTERCHAIN (BY SIMILARITY).  
FT DISULFID 340 356 BY SIMILARITY.  
FT DISULFID 348 417 BY SIMILARITY.  
FT DISULFID 442 516 BY SIMILARITY.  
FT DISULFID 474 490 BY SIMILARITY.  
FT DISULFID 505 534 BY SIMILARITY.  
FT CARBOHYD 149 149 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 481 481 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CONFLICT 260 260 G -> A (IN REF. 1).  
FT CONFLICT 325 325 P -> A (IN REF. 1).  
SQ SEQUENCE 559 AA; 63122 MW; 8CCE2BDB94514D9 CRC64;  
  
Query Match 10.6%; Score 147.5; DB 1; Length 559;  
Best Local Similarity 37.0%; Pred. No. 6.8e-05;  
Matches 34; Conservative 11; Mismatches 34; Indels 13; Gaps 4;  
  
QY 24 CFWDNGHLYREDQTPAPGLRCINWLDAGSLASAPVS-----GAGNHSYCRNPDE 74  
Db 124 CFEEOGITVGTWSTAEGACINW--NSSVLSLKPYNARRPNAIKGLGNGHNYCRNPDR 181  
  
QY 75 DRPGPCVYSGAGVPEKPCEDLRCPETSQ 106  
Db 182 DLK-PWCYVF-XAGKYTFEFCSTPACPKGSE 211  
  
RESULT 6  
HGFA HUMAN STANDARD; PRT; 655 AA.  
AC Q04756; Q14726;  
DT 01-JUN-1994 (Rel. 29, Created)  
DT 01-JUN-1994 (Rel. 29, Last sequence update)  
DT 10-OCT-2003 (Rel. 42, Last annotation update)  
DE Hepatocyte growth factor activator precursor (EC 3.4.21.-) (HGF  
activator) (HGFA).  
GN HGFA.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.  
RC TISSUE=Liver, and Serum;  
RX MEDLINE=93252878; PubMed=7693665;  
RA Miyazawa K., Shimomura T., Kitamura A., Kondo J., Morimoto Y.,  
RA Kitamura N.;  
RT "Molecular cloning and sequence analysis of the cDNA for a human  
serine protease responsible for activation of hepatocyte growth  
factor. Structural similarity of the protease precursor to blood  
coagulation factor XII.";  
RT J. Biol. Chem. 268:10024-10028(1993).  
RL [2]  
RN SEQUENCE OF 40-655 FROM N.A.  
RA Zhao S., Odell C.;  
RL Submitted (FEB-1996) to the EMBL/GenBank/DBJ databases.  
CC -1- FUNCTION: Activates hepatocyte growth factor (HGF) by  
converting it from a single chain to a heterodimeric form.  
CC -1- SUBUNIT: Dimer of a short chain and a long chain linked by a  
disulfide bond.  
CC -1- SUBCELLULAR LOCATION: Secreted as an inactive single-chain  
precursor and is then activated to a heterodimeric form.  
CC -1- TISSUE SPECIFICITY: Liver.  
CC -1- SIMILARITY: Belongs to peptidase family S1.  
CC -1- SIMILARITY: Contains 2 EGF-like domains.  
CC -1- SIMILARITY: Contains 1 fibronectin type I domain.  
CC -1- SIMILARITY: Contains 1 fibronectin type II domain.  
CC -1- SIMILARITY: Contains 1 kringle domain.  
CC -1- CAUTION: It is uncertain whether Met-1 is the initiator.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration

between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (see <http://www.isb-sib.ch/announce/>)  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----  
EMBL; D14012; BAA03113.1; -;  
EMBL; Z69923; -; NOT\_ANNOTATED\_CDS.  
PIR; A46688; A46688.  
HSP; P00763; IDPO.  
MEROPS; S01.228; -;  
Gene; HGNC:4894; HGFAC.  
MIM; 604552; -;  
GO; GO:0005576; C:extracellular; TAS.  
GO; GO:0004252; F:serine-type endopeptidase activity; TAS.  
GO; GO:0006508; P:proteolysis and peptidolysis; TAS.  
InterPro; IPR003003; Cys\_ser\_trypsin.  
InterPro; IPR000742; EGF\_2.  
InterPro; IPR006209; EGF-like.  
InterPro; IPR000083; Fibrctnl.  
InterPro; IPR000562; FN Type II.  
InterPro; IPR006210; IEGF.  
InterPro; IPR000001; Kringle.  
InterPro; IPR01254; Peptidase S1.  
InterPro; IPR001314; Peptidase\_S1A.  
Pfam; PF00008; EGF; 2.  
Pfam; PF00039; fn1; 1.  
Pfam; PF00040; fn2; 1.  
Pfam; PF00051; kringle; 1.  
Pfam; PF00089; trypsin; 1.  
PRINTS; PR00722; CHYMOTRYPSIN.  
PRINTS; PR00013; FNTYPEII.  
PRINTS; PR00018; KRINGLE.  
ProDom; PD00095; FN Type II; 1.  
ProDom; PD000395; Kringle; 1.  
SMART; SM00181; EGF; 2.  
SMART; SM00058; FN1; 1.  
SMART; SM00059; FN2; 1.  
SMART; SM00130; KR; 1.  
SMART; SM00020; Tryp\_Spc; 1.  
PROSITE; PS00022; EGF\_1; 2.  
PROSITE; PS0186; EGF\_2; 1.  
PROSITE; PS00026; EGF\_3; 2.  
PROSITE; PS01253; FIBRONECTIN\_1; 1.  
PROSITE; PS00023; FIBRONECTIN\_2; 1.  
PROSITE; PS00021; KRINGLE\_1; 1.  
PROSITE; PS00070; KRINGLE\_2; 1.  
PROSITE; PS0240; TRYPSIN\_DOM; 1.  
PROSITE; PS00134; TRYPSIN\_HIS; 1.  
PROSITE; PS00135; TRYPSIN\_SER; 1.  
KW Hydrolase; Glycoprotein; Plasma; Serine protease; Kringle; Signal;  
KW EGF-like domain; Repeat; Zymogen.  
FT SIGNAL 1 30  
FT PROPEP 31 372 CLEAVED IN ACTIVE FORM.  
FT CHAIN 373 407 HEPATOCYTE GROWTH FACTOR ACTIVATOR SHORT  
FT CHAIN 408 655 HEPATOCYTE GROWTH FACTOR ACTIVATOR LONG  
FT CHAIN 108 148 FIBRONECTIN TYPE-II.  
FT DOMAIN 160 198 EGF-LIKE 1.  
FT DOMAIN 200 240 FIBRONECTIN TYPE-I.  
FT DOMAIN 241 279 EGF-LIKE 2.  
FT DOMAIN 286 367 KRINGLE.  
FT DOMAIN 408 655 SERINE PROTEASE.  
FT ACT\_SITE 447 447 CHARGE RELAY SYSTEM (BY SIMILARITY).  
FT ACT\_SITE 497 497 CHARGE RELAY SYSTEM (BY SIMILARITY).  
FT ACT\_SITE 598 598 CHARGE RELAY SYSTEM (BY SIMILARITY).  
FT DISULFID 108 133 BY SIMILARITY.  
FT DISULFID 122 148 BY SIMILARITY.  
FT DISULFID 164 175 BY SIMILARITY.  
FT DISULFID 169 186 BY SIMILARITY.  
FT DISULFID 188 197 BY SIMILARITY.

FT DISULFID 202 230 BY SIMILARITY.  
 FT DISULFID 228 237 BY SIMILARITY.  
 FT DISULFID 245 256 BY SIMILARITY.  
 FT DISULFID 250 267 BY SIMILARITY.  
 FT DISULFID 269 278 BY SIMILARITY.  
 FT DISULFID 286 367 BY SIMILARITY.  
 FT DISULFID 307 349 BY SIMILARITY.  
 FT DISULFID 338 362 BY SIMILARITY.  
 FT DISULFID 394 521 INTERCHAIN (BY SIMILARITY).  
 FT DISULFID 432 448 BY SIMILARITY.  
 FT DISULFID 440 510 BY SIMILARITY.  
 FT DISULFID 535 604 BY SIMILARITY.  
 FT DISULFID 567 583 BY SIMILARITY.  
 FT DISULFID 594 622 BY SIMILARITY.  
 FT CARBOHYD 48 48 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 290 290 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 468 468 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 492 492 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 546 546 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CONFLICT 644 644 R -> Q (IN REF. 2).  
 SQ SEQUENCE 655 AA; 2CF72F1E1B862ED7 CRC64;  
 Query Match 10.6%; Score 146.5; DB 1; Length 655;  
 Best Local Similarity 36.9%; Pred. No. 9.9e-05;  
 Matches 41; Conservative 10; Mismatches 41; Indels 19; Gaps 6;  
 QY 24' CFWDNGHLYREDOTSPAPGLRCLNW-----LDAQSLASAPVSGAGNHSYCRNPDED 75  
 DB 286 CFLNGTGTGRGVATSGASGLSCLAWNSDLLYQELHYVDS-VGAAALLGLGPAYACRNPDND 344  
 QY 76 PRGPWCVSGEAGVP---EKPCEDLRCPETTSQALPAFTTEIOB-ASEG 121  
 DB 345 ER-PWCYVVKDSALSWEYCRLEACESL-----TRVQLSPDLLATLPEDAPSG 390  
 RESULT 7  
 ID PLMN HUMAN STANDARD; PRT; 810 AA.  
 AC P00747;  
 DT 21-JUL-1986 (Rel. 01, Created)  
 DT 01-MAR-1989 (Rel. 10, Last sequence update)  
 DT 15-MAR-2004 (Rel. 43, Last annotation update)  
 DE Plasminogen precursor (EC 3.4.21.7) [Contains: Angiostatin].  
 GN PLG.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A., AND VARIANT ASN-472.  
 RX MEDLINE=90202879; PubMed=2318848;  
 RA Petersen T.E., Martzen M.R., Ichinose A., Davie E.W.;  
 RT "Characterization of the gene for human plasminogen, a key proenzyme  
 in the fibrinolytic system.";  
 RL J. Biol. Chem. 265:6104-6111(1990).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=87162490; PubMed=3030813;  
 RA Forsgren M., Raden B., Israelsson M., Larsson K., Heden L.-O.;  
 RT "Molecular cloning and characterization of a full-length cDNA clone  
 for human plasminogen.";  
 RL FEBS Lett. 213:254-260(1987).  
 RN [3]  
 RP SEQUENCE FROM N.A., AND VARIANTS LYS-57; GLN-133; HIS-261; TRP-408;  
 RX ASN-472; VAL-494 AND TRP-523.  
 RA Rieder M.J., Armel T.Z., Carrington D.P., Ozuna M., Kuldanek S.A.,  
 RA Rajkumar N., Toch E.J., Yi Q., Nickerson D.A.;  
 RT Submitted (DEC-2002) to the EMBL/GenBank/DBJ databases.  
 RN [4]  
 RP SEQUENCE OF 20-810, AND VARIANT ASN-472.  
 RA Sottrup-Jensen L., Petersen T.E., Magnusson S.;  
 RL Submitted (JUL-1977) to the PIR data bank.  
 RN [5]

RP SEQUENCE OF 292-810 FROM N.A.  
 RX MEDLINE=85023311; PubMed=6148961;  
 RA Malinowski D.P., Sadler J.E., Davie E.W.;  
 RT "Characterization of a complementary deoxyribonucleic acid coding for  
 human and bovine plasminogen.";  
 RL Biochemistry 23:4243-4250(1984).  
 RN [6]  
 RP SEQUENCE OF 20-100.  
 RX MEDLINE=75093329; PubMed=122932;  
 RA Wiman B., Wallen P.;  
 RT "Structural relationship between 'glutamic acid' and 'lysine' forms  
 of human plasminogen and their interaction with the NH2-terminal  
 activation peptide as studied by affinity chromatography.";  
 RL Eur. J. Biochem. 50:489-494(1975).  
 RN [7]  
 RP SEQUENCE OF 95-580; 581-626; 657-700 AND 732-810, AND VARIANT ASN-472.  
 RX Sottrup-Jensen L., Claeys H., Zajdel M., Petersen T.E., Magnusson S.;  
 RL (In) Davidson J.F., Rawan R.M., Samama M.M., Desnoyers P.C. (eds.);  
 Progress in chemical fibrinolysis and thrombolysis, pp.3:191-209,  
 Raven Press, New York (1978).  
 RN [8]  
 RP SEQUENCE OF 483-604.  
 RX MEDLINE=76043692; PubMed=126863;  
 RA Wiman B., Wallen P.;  
 RT "Amino-acid sequence of the cyanogen-bromide fragment from human  
 plasminogen that forms the linkage between the plasmin chains.";  
 RL Eur. J. Biochem. 58:539-547(1975).  
 RN [9]  
 RP SEQUENCE OF 581-810.  
 RX MEDLINE=77225245; PubMed=142009;  
 RA Wiman B.;  
 RT "Primary structure of the B-chain of human plasmin.";  
 RL Eur. J. Biochem. 76:129-137(1977).  
 RN [10]  
 RP ACTIVE SITE.  
 RX MEDLINE=73149248; PubMed=4694729;  
 RA Robbins K.C., Bernabe P., Arzadon L., Summaria L.;  
 RT "The primary structure of human plasminogen. II. The histidine loop  
 of human plasmin: light (B) chain active center histidine sequence.";  
 RL J. Biol. Chem. 248:1631-1633(1973).  
 RN [11]  
 RP ACTIVE SITE.  
 RX MEDLINE=69234739; PubMed=4240117;  
 RA Groskopf W.R., Summaria L., Robbins K.C.;  
 RT "Studies on the active center of human plasmin. Partial amino acid  
 sequence of a peptide containing the active center serine residue.";  
 RL J. Biol. Chem. 244:3590-3597(1969).  
 RN [12]  
 RP OMEGA-AMINOCARBOXYLIC ACID-BINDING SITES.  
 RX MEDLINE=82213905; PubMed=6919539;  
 RA Trexler M., Vally Z., Paddy L.;  
 RT "Structure of the omega-aminocarboxylic acid-binding sites of human  
 plasminogen. Arginine 70 and aspartic acid 56 are essential for  
 binding of ligand by krigle 4.";  
 RL J. Biol. Chem. 257:7401-7406(1982).  
 RN [13]  
 RP FIBRIN AND OMEGA-AMINOCARBOXYLIC ACID BINDING SITES.  
 RX MEDLINE=85054794; PubMed=6094526;  
 RA Vally Z., Paddy L.;  
 RT "The fibrin-binding site of human plasminogen. Arginines 32 and 34  
 are essential for fibrin affinity of the krigle 1 domain.";  
 RL J. Biol. Chem. 259:13690-13694(1984).  
 RN [14]  
 RP PHOSPHORYLATION SITE SER-597.  
 RX MEDLINE=97345939; PubMed=9201958;  
 RA Wang H., Prock M., Bretthauer R.K., Castellino F.J.;  
 RT "Serine-578 is a major phosphorylation locus in human plasma  
 plasminogen.";  
 RL Biochemistry 36:8100-8106(1997).  
 RN [15]  
 RP CARBOHYDRATE-LINKAGE SITES.  
 RX MEDLINE=88185329; PubMed=3356193;  
 RA Marti T., Schallier J., Rickli E.B., Schmid K., Kamerling J.P.,

RA Gerwig G.J., van Halbeek H., Vliegthart J.F.;  
RT "The N- and O-linked carbohydrate chains of human, bovine and porcine  
RT plasminogen. Species specificity in relation to sialylation and  
RT fucosylation patterns.";  
RL Eur. J. Biochem. 173:57-63 (1988).  
RN [16]  
RP CARBOHYDRATE-LINKAGE SITE SER-268.  
RX MEDLINE=97207306; PubMed=9054441;  
RA Pizzo S.V.;  
RA Pirie-Shepherd S.R., Stevens R.D., Andon N.L., Enghild J.J.,  
RT "Evidence for a novel O-linked sialylated trisaccharide on Ser-248 of  
RT human plasminogen 2.";  
RL J. Biol. Chem. 272:7408-7411 (1997).  
RN [17]  
RP CHARACTERIZATION OF ANGIOSTATIN, AND PARTIAL SEQUENCE.  
RX MEDLINE=95042728; PubMed=7525077;  
RA O'Reilly M.S., Holmgren L., Shing Y., Chen C., Rosenthal R.A.,  
RA Moses M., Iane W.S., Cao Y., Sage E.H., Folkman J.;  
RT "Angiostatin: a novel angiogenesis inhibitor that mediates the  
RT suppression of metastases by a Lewis lung carcinoma.";  
RL Cell 79:315-328 (1994).  
RN [18]  
RP CHARACTERIZATION OF ANGIOSTATIN.  
RX MEDLINE=97238710; PubMed=9102221;  
RA Sim B.K., O'Reilly M.S., Liang H., Fortier A.H., He W., Madsen J.W.,  
RA Lapcevic R., Nacy C.A.;  
RT "A recombinant human angiostatin protein inhibits experimental primary  
RT and metastatic cancer.";  
RL Cancer Res. 57:1329-1334 (1997).  
RN [19]  
RP X-RAY CRYSTALLOGRAPHY (1.9 ANGSTROMS) OF 374-461.  
RX MEDLINE=92031502; PubMed=1657148;  
RA Mulichak A.M., Tulinsky A., Ravichandran K.G.;  
RT "Crystal and molecular structure of human plasminogen kringle 4  
RT refined at 1.9-A resolution.";  
RL Biochemistry 30:10576-10588 (1991).  
RN [20]  
RP X-RAY CRYSTALLOGRAPHY (2.25 ANGSTROMS) OF 374-461.  
RX MEDLINE=92031503; PubMed=1657149;  
RA Wu T.-P., Padmanabhan K., Tulinsky A., Mulichak A.M.;  
RT "The refined structure of the epsilon-aminocaproic acid complex of  
RT human plasminogen kringle 4.";  
RL Biochemistry 30:10589-10594 (1991).  
RN [21]  
RP X-RAY CRYSTALLOGRAPHY (1.67 ANGSTROMS) OF 376-454.  
RA Stec B., Yamano A., Whitlow M., Teeter M.M.;  
RT "Structure of human plasminogen kringle 4 at 1.68 Angstrom and 277 K.  
RT A possible structural role of disordered residues.";  
RL Acta Crystallogr. D 53:169-178 (1997).  
RN [22]  
RP X-RAY CRYSTALLOGRAPHY (2.1 ANGSTROMS) OF 102-181.  
RX MEDLINE=96180681; PubMed=8611560;  
RA Mathews I.J., Vanderhoff-Kanaver P., Castellino F.J., Tulinsky A.;  
RT "Crystal structures of the recombinant kringle 1 domain of human  
RT plasminogen in complexes with the ligands epsilon-aminocaproic acid  
RT and trans-4-(aminomethyl)cyclohexane-1-carboxylic Acid.";  
RL Biochemistry 35:2567-2576 (1996).  
RN [23]  
RP X-RAY CRYSTALLOGRAPHY (1.66 ANGSTROMS) OF 480-563.  
RX MEDLINE=9618034; PubMed=9521645;  
RA Chang Y., Mochalkin I., McCance S.G., Cheng B., Tulinsky A.,  
RA Castellino F.J.;  
RT "Structure and ligand binding determinants of the recombinant kringle  
RT 5 domain of human plasminogen.";  
RL Biochemistry 37:3258-3271 (1998).  
RN [24]  
RP STRUCTURE BY NMR OF 96-184.  
RX MEDLINE=94237157; PubMed=8181475;  
RA Rejante M.R., Llinas M.;  
RT "1H-NMR assignments and secondary structure of human plasminogen  
RT kringle 1.";  
RL Eur. J. Biochem. 221:927-937 (1994).  
RN [25]

RP STRUCTURE BY NMR OF 96-184.  
RX MEDLINE=94237158; PubMed=8181476;  
RA Rejante M.R., Llinas M.;  
RT "Solution structure of the epsilon-aminohexanoic acid complex of  
RT human plasminogen kringle 1.";  
RL Eur. J. Biochem. 221:939-949 (1994).  
RN [26]  
RP STRUCTURE BY NMR OF 183-354.  
RX MEDLINE=96194156; PubMed=8652577;  
RA Soehndel S., Hu C.-K., Marti D., Affolter M., Schaller J., Llinas M.,  
RA Rickli E.E.;  
RT "Recombinant gene expression and 1H NMR characteristics of the  
RT kringle (2 + 3) supermodule: spectroscopic/functional individuality  
RT of plasminogen kringle domains.";  
RL Biochemistry 35:2357-2364 (1996).  
RN [27]  
RP STRUCTURE BY NMR OF 374-461.  
RX MEDLINE=90219023; PubMed=2157850;  
Query Match 10.2%; Score 142; DB 1; Length 810;  
Best Local Similarity 37.2%; Pred. No. 0.0003;  
Matches 32; Conservative 8; Mismatches 34; Indels 12; Gaps 4;  
QY 24 CFWNGHLYREDQTSAPAGRLCLNWLDA-----QSGLASAPVSGAGNHSYCRNPDEPRG 78  
DB 103 CKTGNGNYRTGWSKTKNGITCKQWSSTSPRPFRSPATHPSEGL-EENYCRNPDPNDPOG 161  
QY 79 PWCYVSGEAGVPEKR--PCEDLRCP 102  
DB 162 PWCYTTD---PEKRYDYCDILECEE 183  
RESULT 8  
PLMN\_RAT PLMN\_RAT STANDARD; PRT; 169 AA.  
ID CQ01177;  
DT 01-APR-1993 (Rel. 25, Created)  
DT 01-APR-1993 (Rel. 25, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE Plasminogen (EC 3.4.21.7) (Fragment).  
GN PLG.  
OS Rattus norvegicus (Rat).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
OX NCBI TaxID=10116;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Liver;  
RA MEDLINE=91250378; PubMed=1645711;  
RA Kanalas J.J., Makker S.P.;  
RT "Identification of the rat Heymann nephritis autoantigen (GP330) as a  
RT receptor site for plasminogen.";  
RL J. Biol. Chem. 266:10825-10829 (1991).  
CC -!- FUNCTION: Plasmin dissolves the fibrin of blood clots and acts as  
CC a proteolytic factor in a variety of other processes including  
CC embryonic development, tissue remodeling, tumor invasion, and  
CC inflammation; in ovulation it weakens the walls of the Graafian  
CC follicle. It activates the urokinase-type plasminogen activator,  
CC collagenases and several complement zymogens, such as C1 and C5.  
CC It cleaves fibrin, fibronectin, thrombospondin, laminin and von  
CC Willebrand factor.  
CC -!- CATALYTIC ACTIVITY: Preferential cleavage: Lys-|-Xaa > Arg-|-Xaa;  
CC higher selectivity than trypsin. Converts fibrin into soluble  
CC products.  
CC -!- ENZYME REGULATION: Converted into plasmin by plasminogen  
CC activators, both plasminogen and its activator being bound to  
CC fibrin. Cannot be activated with streptokinase.  
CC -!- MISCELLANEOUS: Plasmin is inactivated by alpha-2-antiplasmin  
CC immediately after dissociation from the clot.  
CC -!- SIMILARITY: Belongs to peptidase family S1. Plasminogen subfamily.  
CC -!- SIMILARITY: Contains 5 kringle domains.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration

```

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; M62832; AAA41884.1; -.
CC PIR; A40522; A40522.
CC HSP; P00747; 1PMK.
CC MEROPS; S01.233; -.
CC InterPro; IPR000001; Kringle.
CC InterPro; IPR001254; Peptidase_S1.
CC InterPro; IPR003966; Peptidase_S1A_pr.
CC Pfam; PF00051; kringle; 2.
CC PRINTS; PR00018; KRINGLE.
CC PRINTS; PR01505; PROTHROMBIN.
CC ProDom; PD000395; Kringle; 2.
CC SMART; SM00130; KR; 1.
CC PROSITE; PS00021; KRINGLE_1; 1.
CC PROSITE; PS00070; KRINGLE_2; 2.
CC PROSITE; PS00240; TRYPSIN_DOM; PARTIAL.
CC PROSITE; PS00134; TRYPSIN_HIS; PARTIAL.
CC PROSITE; PS00135; TRYPSIN_SER; PARTIAL.
CC Hydroxylase; Serine protease; Plasma; Glycoprotein; Fibrinolysis;
KW Tissue remodeling; Blood coagulation; Kringle; Repeat.
FT NON_TER 1
FT DOMAIN <1 10 KRINGLE 3 (BY SIMILARITY).
FT DOMAIN 34 112 KRINGLE 4 (BY SIMILARITY).
FT DOMAIN 139 >169 KRINGLE 5 (BY SIMILARITY).
FT DISULFID 34 112 BY SIMILARITY.
FT DISULFID 55 95 BY SIMILARITY.
FT DISULFID 83 107 BY SIMILARITY.
FT NON_TER 169
FT SEQUENCE 169 AA; 18401 MW; 77A54214C49D010C CRC64;

Query Match 9.9%; Score 138; DB 1; Length 169;
Best Local Similarity 31.8%; Pred. No. 0.00011;
Matches 35; Conservative 13; Mismatches 48; Indels 14; Gaps 5;

QY 24 CFWDNGLHYREDTSPAPGLRCLNWLDA-----QSGLASAPVSGAGNHVYCNRPDEDQG 78
Db 34 CYQNGKSGYRGSTSTNTGKKCSQVSMTHPSHSKTPANFPDQGL-EMMYCRPNPDQDG 92
QY 79 PWCVSGEAGVPEKR--PCEDLRCPETTSQALPAFTTEIQEASGEGGADE 126
Db 93 PWCFTTD----PSVRWEYCNLKCSETGGV--AESAIVPQVPSAPCTSE 136

RESULT 9
UROK_PAPCY STANDARD; PRT; 433 AA.
AC P16227;
DT 01-APR-1990 (Rel. 14, Created)
DT 01-APR-1990 (Rel. 14, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Urokinase-type plasminogen activator precursor (EC 3.4.21.73) (uPA)
DE (U-plasminogen activator).
DE PLAU.
OS Papio cynocephalus (Yellow baboon).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea;
OC Cercopithecoidea; Papio.
OX NCBI_TaxID=9556;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Thoracic aorta;
RX MEDLINE=90287734; PubMed=2113276;
RA Au Y. P. T., Wang T. W., Clowes A. W.;
RT "Nucleotide and deduced amino acid sequences of baboon urokinase-type
RT plasminogen activator."
RL Nucleic Acids Res. 18:3411-3411(1990).
CC -!- CATALYTIC ACTIVITY: Specific cleavage of Arg-|-Val bond in

```

```

CC plasminogen to form plasmin.
CC -!- SUBUNIT: Found in high and low molecular mass forms. Each consists
CC of two chains, A and B. The high molecular mass form contains a
CC long chain A. Cleavage occurs after residue 155 in the low
CC molecular mass form to yield a short A1 chain (By similarity).
CC -!- SIMILARITY: Belongs to peptidase family S1.
CC -!- SIMILARITY: Contains 1 EGF-like domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; X51935; CAA36200.1; -.
CC PIR; S14687; UKBAY.
CC HSP; P00749; 1LMW.
CC MEROPS; S01.231; -.
CC InterPro; IPR009003; Cys_Ser_trypsin.
CC InterPro; IPR006209; EGF_Like.
CC InterPro; IPR006210; EGF.
CC InterPro; IPR000001; Kringle.
CC InterPro; IPR008293; Pept_S1A_uPA.
CC InterPro; IPR001254; Peptidase_S1.
CC InterPro; IPR001314; Peptidase_S1A.
CC Pfam; PF00051; kringle; 1.
CC Pfam; PF00089; trypsin; 1.
CC PIRSF; PIRSF001144; Uro_plasm_act; 1.
CC PRINTS; PR00722; CHYMOTRYPSIN.
CC PRINTS; PR00018; KRINGLE.
CC ProDom; PD000395; Kringle; 1.
CC SMART; SM00181; EGF; 1.
CC SMART; SM00130; KR; 1.
CC SMART; SM00020; Tryp_Spc; 1.
CC PROSITE; PS00022; EGF_1; 1.
CC PROSITE; PS01186; EGF_2; FALSE_NEG.
CC PROSITE; PS00026; EGF_3; 1.
CC PROSITE; PS00021; KRINGLE_1; 1.
CC PROSITE; PS00070; KRINGLE_2; 1.
CC PROSITE; PS00240; TRYPSIN_DOM; 1.
CC PROSITE; PS00134; TRYPSIN_HIS; 1.
CC PROSITE; PS00135; TRYPSIN_SER; 1.
CC Plasminogen activation; Hydrolase; Serine protease; Glycoprotein;
KW Kringle; EGF-like domain; Zymogen; Signal.
FT SIGNAL 1 20
FT CHAIN 21 433 UROKINASE-TYPE PLASMINOGEN ACTIVATOR.
FT CHAIN 21 176 CHAIN A (BY SIMILARITY).
FT CHAIN 155 176 SHORT A CHAIN (A1) (BY SIMILARITY).
FT CHAIN 178 433 CHAIN B (BY SIMILARITY).
FT DOMAIN 26 62 EGF-LIKE.
FT DOMAIN 69 150 KRINGLE.
FT DOMAIN 151 177 CONNECTING PEPTIDE.
FT DOMAIN 178 433 SERINE PROTEASE.
FT DISULFID 30 38 BY SIMILARITY.
FT DISULFID 32 50 BY SIMILARITY.
FT DISULFID 52 61 BY SIMILARITY.
FT DISULFID 167 298 INTERCHAIN (BY SIMILARITY).
FT DISULFID 208 224 BY SIMILARITY.
FT DISULFID 216 287 BY SIMILARITY.
FT DISULFID 315 384 BY SIMILARITY.
FT DISULFID 347 363 BY SIMILARITY.
FT DISULFID 374 402 BY SIMILARITY.
FT ACT_SITE 223 CHARGE RELAY SYSTEM.
FT ACT_SITE 274 CHARGE RELAY SYSTEM.
FT ACT_SITE 378 CHARGE RELAY SYSTEM.
FT CARBOHYD 324 324 N-LINKED (GLCNAC...) (BY SIMILARITY).
FT SEQUENCE 433 AA; 48595 MW; 816D22DFEDDC8792 CRC64;

Query Match 9.9%; Score 138; DB 1; Length 433;
Best Local Similarity 32.7%; Pred. No. 0.00032;

```

Matches 32; Conservative 14; Mismatches 36; Indels 16; Gaps 4;

QY 24 CFWDNGHLYREDQTSFAPGLRCNLWDA-----QSGLASAPVSGAGNHSYCRNPDPDP 76

Db 69 CVEGNGHYFVRGKASDTMGSRCLAWNSATVLTQTYHAHRSDALQLGLGKHNYCRNPD-NR 127

QY 77 RGPWCYVGEAGVPEK-----RPGCDLRCPTTSQAL 108

Db 128 RRPWCYV--QVGLKORVQECWVHNCADGKXSPSPPEEL 163

RESULT 10

UOK\_PIG

ID UOK\_PIG STANDARD; PRT; 442 AA.

AC P04185;

DT 20-MAR-1987 (Rel. 04, Created)

DT 13-AUG-1987 (Rel. 05, Last sequence update)

DT 28-FEB-2003 (Rel. 41, Last annotation update)

DE Urokinase-type plasminogen activator precursor (EC 3.4.21.73) (uPA)

DE (U-plasminogen activator).

GN PLAU.

OS Sus scrofa (Pig).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Euthera; Cetartiodactyla; Suina; Suidae; Sus.

OX NCBI\_TaxID=9823;

RN [1]

RP SEQUENCE FROM N.A.

RC TISSUE=Kidney;

RX MEDLINE=85387954; PubMed=6056832;

RA Nagamine Y., Pearson D., Alcus M.S., Reich E.;

RT "cDNA and gene nucleotide sequence of porcine plasminogen activator.";

RL Nucleic Acids Res. 12:9525-9541(1984).

RN [2]

RP REVISION TO 241.

RA Nagamine Y.

RL Submitted (DEC-1986) to the PIR data bank.

CC -!- CATALYTIC ACTIVITY: Specific cleavage of Arg-|-Val bond in plasminogen to form plasmin.

CC -!- SIMILARITY: Belongs to peptidase family S1.

CC -!- SIMILARITY: Contains 1 EGF-like domain.

CC -!- SIMILARITY: Contains 1 kringle domain.

CC [2]

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC [2]

DR EMBL; X01648; CAA25806.1; -;

DR EMBL; X02724; CAA26511.1; -;

DR PIR; A00932; URPG.

DR HSSP; P00749; 1KDU.

DR MEROPS; S01.231; -;

DR InterPro; IPR009003; Cys Ser trypsin.

DR InterPro; IPR006203; EGF-like.

DR InterPro; IPR000001; Kringle.

DR InterPro; IPR008293; Pept\_S1a\_uPA.

DR InterPro; IPR001254; Peptidase\_S1.

DR InterPro; IPR001314; Peptidase\_S1A.

DR Pfam; PF000051; Kringle; 1.

DR Pfam; PF00089; trypsin; 1.

DR PIRSF; PIRSF001144; Uro plas act; 1.

DR PRINTS; PR00722; CHYMOTRYPSIN.

DR PRINTS; PR00018; KRINGLE.

DR ProDom; PD000395; Kringle; 1.

DR SMART; SM00130; KR; 1.

DR SMART; SM00020; Tryp SPC; 1.

DR PROSITE; PS00022; EGF\_1; 1.

DR PROSITE; PS01186; EGF\_2; FALSE\_NEG.

DR PROSITE; PS50026; EGF\_3; 1.

DR PROSITE; PS00021; KRINGLE\_1; 1.

DR PROSITE; PS50070; KRINGLE 2; 1.

DR PROSITE; PS00240; TRYPSIN\_DOM; 1.

DR PROSITE; PS00134; TRYPSIN\_HLS; 1.

DR PROSITE; PS00135; TRYPSIN\_SER; 1.

KW Plasminogen activation; Hydrolase; Serine protease; Glycoprotein;

KW Kringle; EGF-like domain; Zymogen; Signal.

KW SIGNAL 1 20 BY SIMILARITY.

FT CHAIN 21 442 UROKINASE-TYPE PLASMINOGEN ACTIVATOR.

FT CHAIN 21 188 CHAIN A (BY SIMILARITY).

FT CHAIN 190 442 CHAIN B (BY SIMILARITY).

FT CHAIN 29 65 EGF-LIKE.

FT DOMAIN 72 153 KRINGLE.

FT DOMAIN 154 189 CONNECTING PEPTIDE.

FT DOMAIN 190 442 SERINE PROTEASE.

FT CARBOHYD 152 152 N-LINKED (GLCNAC. . .).

FT DISULFID 33 41 BY SIMILARITY.

FT DISULFID 35 53 BY SIMILARITY.

FT DISULFID 55 64 BY SIMILARITY.

FT DISULFID 179 310 INTERCHAIN (BY SIMILARITY).

FT DISULFID 220 236 BY SIMILARITY.

FT DISULFID 228 299 BY SIMILARITY.

FT DISULFID 324 393 BY SIMILARITY.

FT DISULFID 356 372 BY SIMILARITY.

FT DISULFID 383 411 BY SIMILARITY.

FT ACT\_SITE 235 235 CHARGE RELAY SYSTEM.

FT ACT\_SITE 286 286 CHARGE RELAY SYSTEM.

FT ACT\_SITE 387 387 CHARGE RELAY SYSTEM.

FT CONFLICT 241 241 Q -> H (IN REF. 1; CAA25806).

FT CONFLICT 242 242 Q -> H (IN REF. 1; CAA26511).

FT CONFLICT 288 288 A -> GS (IN REF. 1; CAA25806).

SQ SEQUENCE 442 AA; 49116 MW; EE32FCBF501321EE CRC64;

Query Match 9.8%; Score 135.5; DB 1; Length 442;

Best Local Similarity 36.9%; Pred. No. 0.00054;

Matches 38; Conservative 7; Mismatches 37; Indels 21; Gaps 5;

QY 24 CFWDNGHLYREDQTSFAPGLRCNLWDAQSGL-----ASAPVS---GAGNHSYCRNPDPDP 76

Db 72 CFEENGHSYRGKANTNTGGRPCLPWNSATVLTQTYHAHRSDALQLGLGKHNYCRNPD-NQ 130

QY 77 RGPWCYV-----GEAGVP-----EKPCDRLRCPTTSQ 106

Db 131 RRPWCYVQVGLKQVQCVWVPCNGSGSHRFPAYDGNPFSTPE 173

RESULT 11

URTB\_DESRO

ID URTB\_DESRO STANDARD; PRT; 431 AA.

AC P98121;

DT 01-FEB-1996 (Rel. 33, Created)

DT 01-FEB-1996 (Rel. 33, Last sequence update)

DT 28-FEB-2003 (Rel. 41, Last annotation update)

DE Salivary plasminogen activator beta precursor (EC 3.4.21.68) (DSPA beta).

DE beta).

OS Desmodus rotundus (Vampire bat).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Euthera; Chiroptera; Microchiroptera; Phyllostomidae;

OC Desmodontinae; Desmodus.

OX NCBI\_TaxID=9430;

RN [1]

RP SEQUENCE FROM N.A.

RC TISSUE=Salivary gland;

RX MEDLINE=92039036; PubMed=937019;

RA Kraetzschmar J., Haendler B., Langer G., Boidol W., Bringmann P.,

RA Alagon A., Donner P., Schleuning W.D.;

RT "The plasminogen activator family from the salivary gland of the vampire bat Desmodus rotundus: cloning and expression.";

RL Gene 105:229-237(1991).

RN [2]

RN CHARACTERIZATION.

RP MEDLINE=9339059; PubMed=1309059;

RX Schleuning W.-D., Alagon A., Boidol W., Bringmann P., Petri T.,

RA Kraetzschmar J., Haendler B., Langer G., Baldus B., Witt W.,

RA Donner P.;  
 RT "plasminogen activators from the saliva of Desmodus rotundus (common vampire bat): unique fibrin specificity.";  
 RL Ann. N.Y. Acad. Sci. 667:395-403(1992).  
 CC -|- FUNCTION: Probably essential to support the feeding habits of this exclusively haematophagous animal. Probable potent thrombolytic agent.  
 CC -|- CATALYTIC ACTIVITY: Specific cleavage of Arg-|-Val bond in plasminogen to form plasmin.  
 CC -|- SUBUNIT: Monomer.  
 CC -|- SIMILARITY: Belongs to peptidase family S1.  
 CC -|- SIMILARITY: Contains 1 EGF-like domain.  
 CC -|- SIMILARITY: Contains 1 kringle domain.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation at the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: M63989; AAA1594.1; --  
 CC PIR: JS0599; JS0599.  
 CC HSP: P98119; IA51.  
 CC MEROPS: S01.239; -- Cys\_Ser trypsin.  
 CC InterPro: IPR009003; --  
 CC InterPro: IPR006209; EGF-like.  
 CC InterPro: IPR006210; IEGF.  
 CC InterPro: IPR000001; Kringle.  
 CC InterPro: IPR001254; Peptidase\_S1.  
 CC InterPro: IPR001314; Peptidase\_S1A.  
 CC Pfam: PF00008; EGF; 1.  
 CC Pfam: PF00051; kringle; 1.  
 CC Pfam: PF00089; trypsin; 1.  
 CC PRINTS: PR00722; CHYMOTRYPSIN.  
 CC PRINTS: PR00018; KRINGLE.  
 CC ProDom: PD000395; Kringle; 1.  
 CC SMART: SM00181; EGF; 1.  
 CC SMART: SM00130; KR; 1.  
 CC SMART: SM00020; Tryp\_Spc; 1.  
 CC PROSITE: PS00022; EGF\_1; 1.  
 CC PROSITE: PS01186; EGF\_2; 1.  
 CC PROSITE: PS00026; EGF\_3; 1.  
 CC PROSITE: PS00021; KRINGLE\_1; 1.  
 CC PROSITE: PS00070; KRINGLE\_2; 1.  
 CC PROSITE: PS00240; TRYPSIN\_DOM; 1.  
 CC PROSITE: PS00134; TRYPSIN\_HIS; 1.  
 CC PROSITE: PS00135; TRYPSIN\_SER; 1.  
 CC Plasminogen activation; Hydrolase; Serine protease; Glycoprotein; Kringle; EGF-like domain; Signal; Multigene family.  
 CC SIGNAL 36  
 CC CHAIN 37 431 SALIVARY PLASMINOGEN ACTIVATOR BETA.  
 CC DOMAIN 37 75 EGF-LIKE.  
 CC DOMAIN 82 163 KRINGLE.  
 CC DOMAIN 179 431 SERINE PROTEASE.  
 CC ACT\_SITE 226 226 CHARGE RELAY SYSTEM (BY SIMILARITY).  
 CC ACT\_SITE 275 275 CHARGE RELAY SYSTEM (BY SIMILARITY).  
 CC ACT\_SITE 382 382 CHARGE RELAY SYSTEM (BY SIMILARITY).  
 CC ACT\_SITE 41 52 BY SIMILARITY.  
 CC DISULFID 46 63 BY SIMILARITY.  
 CC DISULFID 65 74 BY SIMILARITY.  
 CC DISULFID 82 163 BY SIMILARITY.  
 CC DISULFID 103 145 BY SIMILARITY.  
 CC DISULFID 134 158 BY SIMILARITY.  
 CC DISULFID 168 299 BY SIMILARITY.  
 CC DISULFID 211 227 BY SIMILARITY.  
 CC DISULFID 219 288 BY SIMILARITY.  
 CC DISULFID 313 388 BY SIMILARITY.  
 CC DISULFID 345 361 BY SIMILARITY.  
 CC DISULFID 378 406 BY SIMILARITY.  
 CC CARBOHYD 139 139 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 CC CARBOHYD 352 352 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 CC

SQ SEQUENCE 431 AA; 48221 MW; 699B5E675B162CBF CRC64;  
 Query Match 9.7%; Score 134; DB 1; Length 431;  
 Best Local Similarity 38.6%; Pred. No. 0.0007;  
 Matches 27; Conservative 9; Mismatches 20; Indels 14; Gaps 3;  
 QY 24 CFWDNGHLYREDQTSAPGLRCLNWLDAOSGL-----ASAPVSGAGNHSVCENPD 73  
 DB 82 CYDQGVYRGWTSTESGAQCINW---NSNLTTRTYNGRRSDATLGLGNHNYCRNPD 138  
 QY 74 EDPRGPWCYV 83  
 DB 139 NNSK-PWCYV 147  
 RESULT 12  
 ID\_URT2 DESRO STANDARD; PRT; 477 AA.  
 AC P15638;  
 DT 01-APR-1990 (Rel. 14, Created)  
 DT 01-FEB-1996 (Rel. 33, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Salivary plasminogen activator alpha 2 precursor (EC 3.4.21.68) (DSPA alpha-2) (BAT-PA) (T-plasminogen activator).  
 OS Desmodus rotundus (Vampire bat).  
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; CC Mammalia; Eutheria; Chiroptera; Microchiroptera; Phyllostomidae; CC Desmodontinae; Desmodus.  
 CC NCBI\_TaxID=9430;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Salivary gland;  
 RX MEDLINE=92039036; PubMed=1937019;  
 RA Kraetzschmar J., Haendler B., Langer G., Boidel W., Bringmann P., Alagon A., Donner P., Schlemming W.D.;  
 RT "The plasminogen activator family from the salivary gland of the vampire bat Desmodus rotundus: cloning and expression.";  
 RL Gene 105:229-237(1991).  
 RN [2]  
 RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.  
 RC TISSUE=Salivary gland;  
 RX MEDLINE=90036867; PubMed=2509450;  
 RA Gardell S.J., Duong L.T., Diehl R.E., York J.D., Hare T.R., Register R.B., Jacobs J.W., Dixon R.A.F., Friedman P.A.;  
 RT "Isolation, characterization, and cDNA cloning of a vampire bat salivary plasminogen activator.";  
 RL J. Biol. Chem. 264:17947-17952(1989).  
 RN [3]  
 RP CHARACTERIZATION.  
 RX MEDLINE=9339059; PubMed=1309059;  
 RA Schlemming W.-D., Alagon A., Boidel W., Bringmann P., Petri T., Kraetzschmar J., Haendler B., Langer G., Baldus B., Witt W., Donner P.;  
 RT "Plasminogen activators from the saliva of Desmodus rotundus (common vampire bat): unique fibrin specificity.";  
 RL Ann. N.Y. Acad. Sci. 667:395-403(1992).  
 CC -|- FUNCTION: Probably essential to support the feeding habits of this exclusively haematophagous animal. Probable potent thrombolytic agent.  
 CC -|- CATALYTIC ACTIVITY: Specific cleavage of Arg-|-Val bond in plasminogen to form plasmin.  
 CC -|- ENZYME REGULATION: Activity toward plasminogen is stimulated in the presence of fibrin I.  
 CC -|- SUBUNIT: Monomer.  
 CC -|- DOMAIN: The fibronectin type-I domain mediates binding to fibrin, and the kringle domain apparently mediates fibrin-induced stimulation of activity.  
 CC -|- SIMILARITY: Belongs to peptidase family S1.  
 CC -|- SIMILARITY: Contains 1 EGF-like domain.  
 CC -|- SIMILARITY: Contains 1 fibronectin type I domain.  
 CC -|- SIMILARITY: Contains 1 kringle domain.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration





DR PRINTS; PR00722; CHYMOTRYPSIN.  
 DR PRINTS; PR00018; KRINGLE.  
 DR PRINTS; PR01505; PROTHROMBIN.  
 DR PR0001; PD000395; Kringle; 4.  
 DR SMART; SM00130; KR; 4.  
 DR SMART; SM00473; PAN\_AP; 1.  
 DR SMART; SM00020; TRYD\_SPC; 1.  
 DR PROSITE; PS00021; KRINGLE 1; 4.  
 DR PROSITE; PS00070; KRINGLE 2; 4.  
 DR PROSITE; PS00240; TRYPsin; 1.  
 DR PROSITE; PS00240; TRYPsin; 1.  
 KW Kringle; Glycoprotein; Serine protease homolog; Repeat; Signal.  
 FT SIGNAL 1 18  
 FT CHAIN 13 716  
 FT DOMAIN 19 109 PAP.  
 FT DOMAIN 19 109 KRINGLE 1.  
 FT DOMAIN 110 186 KRINGLE 2.  
 FT DOMAIN 191 268 KRINGLE 3.  
 FT DOMAIN 292 370 KRINGLE 4.  
 FT DOMAIN 379 457 KRINGLE 4.  
 FT DOMAIN 489 716 SERINE PROTEASE-LIKE.  
 FT DISULFID 56 78 BY SIMILARITY.  
 FT DISULFID 60 66 BY SIMILARITY.  
 FT DISULFID 110 186 BY SIMILARITY.  
 FT DISULFID 131 169 BY SIMILARITY.  
 FT DISULFID 157 181 BY SIMILARITY.  
 FT DISULFID 191 268 BY SIMILARITY.  
 FT DISULFID 194 333 INTERCHAIN (BY SIMILARITY).  
 FT DISULFID 212 251 BY SIMILARITY.  
 FT DISULFID 240 263 BY SIMILARITY.  
 FT DISULFID 292 370 BY SIMILARITY.  
 FT DISULFID 313 352 BY SIMILARITY.  
 FT DISULFID 341 364 BY SIMILARITY.  
 FT DISULFID 379 457 BY SIMILARITY.  
 FT DISULFID 400 440 BY SIMILARITY.  
 FT DISULFID 428 452 BY SIMILARITY.  
 FT DISULFID 477 593 INTERCHAIN (BY SIMILARITY).  
 FT DISULFID 512 528 BY SIMILARITY.  
 FT DISULFID 607 672 BY SIMILARITY.  
 FT DISULFID 637 651 BY SIMILARITY.  
 FT DISULFID 662 690 BY SIMILARITY.  
 FT CARBOHYD 72 72 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 173 173 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 305 305 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 620 620 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CONFLICT 19 19 P -> Q (IN GENOMIC SEQUENCE).  
 SQ SEQUENCE 716 AA; 80588 MW; BCC02F89213ACC CRC64;  
  
 Query Match 9.7%; Score 134; DB 1; Length 716;  
 Best Local Similarity 28.2%; Pred. No. 0.0012;  
 Matches 42; Conservative 9; Mismatches 46; Indels 52; Gaps 7;  
  
 QY 1 LLAVQAPLVSNMLLAAYGSG-----GCFWNGHLYREDQTSAPGLRCLNW 48  
 DB 80 LLPWTQ-----HSLHTQYHSLCHLFQKKDYVTCIMDNGVSYRGTVARTAGGLPCQAW 134  
 QY 49 ---LDAQSGLASAPVSGAGNHSVCRPDDEPRGFWCVS----- 84  
 DB 135 SRFPNDHKYTPTPKNGL-BENFCRNPDPGPRGWCYTTNRSVRFQSGIKTCREAVCVL 193  
 QY 85 -----GEAGVPEK-RPCE--DLRCPET 103  
 DB 194 CNGEDYRGVEVDVTSRGECQWDLQPHS 222  
  
 RESULT 14  
 URTG DESRO STANDARD; PRT; 394 AA.  
 AC P49150;  
 DT 01-FEB-1996 (Rel. 33, Created)  
 DT 01-FEB-1996 (Rel. 33, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Salivary plasminogen activator gamma precursor (EC 3.4.21.68) (DSFA  
 gamma).  
 OS Desmodus rotundus (Vampire bat).

OC Bukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Chiroptera; Microchiroptera; Phyllostomidae;  
 OC Desmodontinae; Desmodus.  
 OX NCBI\_TaxID=9430;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Salivary gland;  
 RX MEDLINE=92039036; PubMed=1937019;  
 RA Kraetzschmar J., Haendler B., Langer G., Boidol W., Bringmann P.,  
 Alagon A., Donner P., Schleuning W.D.;  
 RT "The plasminogen activator family from the salivary gland of the  
 vampire bat Desmodus rotundus: cloning and expression.";  
 RL Gene 105:229-237 (1991).  
 RN [2]  
 RP CHARACTERIZATION.  
 RX MEDLINE=93393059; PubMed=1309059;  
 RA Schleuning W.-D., Alagon A., Boidol W., Bringmann P., Petri T.,  
 Kraetzschmar J., Haendler B., Langer G., Baldus B., Witt W.,  
 Donner P.;  
 RT "Plasminogen activators from the saliva of Desmodus rotundus (common  
 vampire bat): unique fibrin specificity.";  
 RL Ann. N.Y. Acad. Sci. 667:395-403 (1992).  
 CC -!- FUNCTION: Probably essential to support the feeding habits of this  
 exclusively haematophagous animal. Probable potent thrombolytic  
 agent.  
 CC -!- CATALYTIC ACTIVITY: Specific cleavage of Arg-|-Val bond in  
 plasminogen to form plasmin.  
 CC -!- SUBUNIT: Monomer.  
 CC -!- SIMILARITY: Belongs to peptidase family S1.  
 CC -!- SIMILARITY: Contains 1 kringle domain.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL Outstation  
 at the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC ENBL; M63990; AAA31595.1; -  
 DR PIR; J06000; J06000.  
 DR HSP; P98119; IAS1.  
 DR MEROPS; S01.239; -  
 DR InterPro; IPR009003; Cys Ser trypsin.  
 DR InterPro; IPR000001; Kringle\_S1.  
 DR InterPro; IPR001254; Peptidase\_S1.  
 DR InterPro; IPR001314; Peptidase\_S1A.  
 DR Pfam; PF00051; kringle; 1.  
 DR Pfam; PF00089; trypsin; 1.  
 DR PRINTS; PR00722; CHYMOTRYPSIN.  
 DR PRINTS; PR00018; KRINGLE.  
 DR ProDom; PD000395; Kringle; 1.  
 DR SMART; SM00130; KR; 1.  
 DR SMART; SM00020; TRYD\_SPC; 1.  
 DR PROSITE; PS00021; KRINGLE\_1; 1.  
 DR PROSITE; PS00070; KRINGLE\_2; 1.  
 DR PROSITE; PS00240; TRYPsin; 1.  
 DR PROSITE; PS00134; TRYPsin; 1.  
 DR PROSITE; PS00135; TRYPsin; 1.  
 DR Plasminogen activator; Hydrolase; Serine protease; Glycoprotein;  
 Kringle; Signal; Multigene family.  
 FT SIGNAL 1 36 POTENTIAL.  
 FT CHAIN 37 394 SALIVARY PLASMINOGEN ACTIVATOR GAMMA.  
 FT DOMAIN 45 126 KRINGLE.  
 FT DOMAIN 142 394 SERINE PROTEASE.  
 FT ACT\_SITE 189 189 CHARGE RELAY SYSTEM (BY SIMILARITY).  
 FT ACT\_SITE 238 238 CHARGE RELAY SYSTEM (BY SIMILARITY).  
 FT ACT\_SITE 345 345 CHARGE RELAY SYSTEM (BY SIMILARITY).  
 FT DISULFID 45 126 BY SIMILARITY.  
 FT DISULFID 65 108 BY SIMILARITY.  
 FT DISULFID 97 121 BY SIMILARITY.  
 FT DISULFID 131 262 BY SIMILARITY.  
 FT DISULFID 174 190 BY SIMILARITY.



RP STRUCTURE BY NMR OF 67-155.  
RX MEDLINE=93003110; PubMed=1327118;  
RA Li X., Smith R.A.G., Dobson C.M.;  
RT "Sequential 1H NMR assignments and secondary structure of the kringle  
RL domain from urokinase.";  
RN Biochemistry 31:9562-9571(1992).  
[15]  
RP STRUCTURE BY NMR OF 67-155.  
RX MEDLINE=94149701; PubMed=8107091;  
RA Li X., Bokman A.M., Llinas M., Smith R.A.G., Dobson C.M.;  
RT "Solution structure of the kringle domain from urokinase-type  
RL plasminogen activator.";  
RN J. Mol. Biol. 235:1548-1559(1994).  
[16]  
RP VARIANT LEU-141.  
RX MEDLINE=96186279; PubMed=8652631;  
RA Yoshimoto M., Ushiyama Y., Sakai M., Tamaki S., Hara H., Takahashi K.,  
Sawasaki Y., Hanada K.;  
RT "Characterization of single chain urokinase-type plasminogen  
RL activator with a novel amino-acid substitution in the kringle  
RN structure.";  
RN Biochim. Biophys. Acta 1293:83-89(1996).  
[17]  
RP VARIANT LEU-141.  
RX MEDLINE=97218551; PubMed=9065988;  
RA Conne B., Berzy M., Belin D.;  
RT "Detection of polymorphisms in the human urokinase-type plasminogen  
RL activator gene.";  
RN Thromb. Haemost. 77:434-435(1997).  
[18]  
RP ERRATUM.  
RA Conne B., Berzy M., Belin D.;  
RL Thromb. Haemost. 78:973-973(1997).  
[19]  
RP VARIANT LEU-141.  
RX MEDLINE=97337920; PubMed=9194591;  
RA Turkmen B., Schmitt M., Schmalfeldt B., Trommler P., Hell W.,  
Creutzburg S., Graeff H., Magdolen V.;  
RT "Mutational analysis of the genes encoding urokinase-type plasminogen  
RL activator (uPA) and its inhibitor PAI-1 in advanced ovarian cancer.";  
RN Electrophoresis 18:686-689(1997).  
CC -|- FUNCTION: Potent plasminogen activator and is clinically used for  
CC therapy of thrombolytic disorders.  
CC -|- CATALYTIC ACTIVITY: Specific cleavage of Arg-|-Val bond in  
CC plasminogen to form plasmin.  
CC -|- SUBUNIT: Found in high and low molecular mass forms. Each consists  
CC of two chains, A and B. The high molecular mass form contains a  
CC long chain A. Cleavage occurs after residue 155 in the low  
CC molecular mass form to yield a short A1 chain.  
CC -|- PHARMACEUTICAL: Available under the name Abbokinase (Abbott). Used  
CC in Pulmonary Embolism (PE) to initiate fibrinolysis.  
CC -|- SIMILARITY: Belongs to peptidase family S1.  
CC -|- SIMILARITY: Contains 1 EGF-like domain.  
CC -|- SIMILARITY: Contains 1 kringle domain.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL; X02419; CAA26268.1; -  
CC EMBL; M15476; AAA61253.1; -  
CC EMBL; D00244; BAA00175.1; -  
CC EMBL; D11143; BAA01919.1; -  
CC EMBL; X02760; CAA26535.1; -  
CC EMBL; AF377330; AAK53822.1; -  
CC EMBL; BC013575; AAI13575.1; -  
CC EMBL; K03226; AAC97138.1; -  
CC EMBL; K02286; AAA61252.1; -  
CC EMBL; A21571; CAA01559.1; -

DR EMBL; A18397; CAA01390.1; -  
DR PIR; A00931; UKHU.  
DR PDB; 1KDU; 31-OCT-93.  
Query Match 9.5%; Score 132; DB 1; Length 431;  
Best Local Similarity 32.7%; Pred. No. 0.001; 33; Indels 22; Gaps 5;  
Matches 33; Conservative 13; Mismatches 13;  
QY 24 CFWDNGHLYREDQTSAPGLRCINWIDA-----QSGLASAPVSGAGNHSYCRNPDED 76  
Db 70 CYEGNGHYRGKASDTMGRPCLPWSATVLQQTVAHRSALQGLGKHNYCRNPD-NR 128  
QY 77 RGPWCYVSGEAGVPEKRP-----CEDLRCPETTSQAL 108  
Db 129 RRPWCYV--QVGL---KPLVQECMVHDCADGKKPSPPEEL 164  
Search completed: March 17, 2004, 07:04:24  
Job time : 11.5787 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model  
Run on: March 17, 2004, 06:59:03 ; Search time 47.9569 Seconds  
(without alignments)  
1723.750 Million cell updates/sec

Title: US-10-057-951-2\_COPY\_2\_263  
Perfect score: 1388  
Sequence: 1 LLAWQAFVLSNMLLAAYG.....PVDPQEGSTPLMGQATFGA 262

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues  
Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

- Database : SPTREMBL.25.\*
- 1: sp\_archaea.\*
  - 2: sp\_bacteria.\*
  - 3: sp\_fungi.\*
  - 4: sp\_human.\*
  - 5: sp\_invertebrate.\*
  - 6: sp\_mammal.\*
  - 7: sp\_mhc.\*
  - 8: sp\_organelle.\*
  - 9: sp\_phage.\*
  - 10: sp\_plant.\*
  - 11: sp\_rodent.\*
  - 12: sp\_virus.\*
  - 13: sp\_vertebrate.\*
  - 14: sp\_unclassified.\*
  - 15: sp\_virus.\*
  - 16: sp\_bacterioph.\*
  - 17: sp\_archaeap.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1388	100.0	263	4 Q96FE7	Q96fe7 homo sapien
2	1385	99.8	263	4 Q00318	Q00318 homo sapien
3	1374	99.0	263	4 Q8NCJ9	Q8ncj9 homo sapien
4	1115.5	80.4	264	11 Q7TJW8	Q7tjw8 mus musculus
5	1111.5	80.1	264	11 Q811Z2	Q811z2 mus musculus
6	1034	74.5	234	4 Q86YW2	Q86yw2 homo sapien
7	821.5	59.2	213	11 Q811Z3	Q811z3 mus musculus
8	611	44.0	263	13 Q7SXB3	Q7sxb3 brachydanio
9	167	12.0	562	6 Q8SQ23	Q8sq23 sus scrofa
10	154.5	11.1	291	4 Q7Z7N2	Q7z7n2 homo sapien
11	154.5	11.1	516	4 Q9BU99	Q9bux9 homo sapien
12	154.5	11.1	562	4 Q86YK8	Q86yk8 homo sapien
13	150	10.8	653	11 Q8VCS4	Q8vcs4 mus musculus
14	145.5	10.5	564	6 Q8MKB1	Q8mbk1 oryctolagus
15	142	10.2	810	4 Q15146	Q15146 homo sapien
16	139.5	10.1	385	5 Q25101	Q25101 herdmania m

17	137	9.9	391	5 Q86PQ9	Q86pq9 cryptospori
18	137	9.9	716	11 P70521	P70521 rattus norv
19	136.5	9.8	420	13 Q90504	Q90504 eptaretus
20	136	9.8	421	13 Q8AXX3	Q8axx3 xenopus lae
21	135	9.7	90	4 Q8NG20	Q8ng20 homo sapien
22	135	9.7	812	11 Q8R0W3	Q8r0w3 rattus norv
23	134.5	9.7	395	4 Q9BZW1	Q9bzw1 homo sapien
24	134.5	9.7	704	13 Q90865	Q90865 gallus gall
25	134	9.7	716	11 Q91XG8	Q91xg8 mus musculu
26	131.5	9.5	313	13 Q9PU78	Q9pu78 crocodylus
27	131	9.4	154	4 Q96SE8	Q96se8 homo sapien
28	131	9.4	608	13 Q9PTW7	Q9ptw7 struthio ca
29	130.5	9.4	709	13 Q7ZTN9	Q7ztn9 xenopus lae
30	130.5	9.4	717	13 P70006	P70006 xenopus lae
31	129	9.3	616	6 Q97507	Q97507 sus scrofa
32	128.5	9.3	806	6 Q18783	Q18783 macropus eu
33	127	9.1	157	6 Q9TVA8	Q9tva8 bos taurus
34	126.5	9.1	433	6 Q8MIL0	Q8mil0 oryctolagus
35	126.5	9.1	433	6 Q8MHY7	Q8mhy7 oryctolagus
36	126.5	9.1	728	11 Q8C9G5	Q8c9g5 mus musculu
37	126	9.1	716	13 Q91691	Q91691 xenopus lae
38	125.5	9.0	560	4 Q14520	Q14520 homo sapien
39	125	9.0	622	4 Q7Z7P3	Q7z7p3 homo sapien
40	125	9.0	730	6 Q867B7	Q867b7 canis fami
41	124	8.9	429	13 Q8AVB0	Q8avb0 brachydanio
42	123	8.9	728	6 Q9BH09	Q9bh09 felis silve
43	120	8.6	231	11 Q8C6L2	Q8c6l2 mus musculu
44	120	8.6	612	13 Q804W7	Q804w7 fugu rubrip
45	119	8.6	334	6 Q46507	Q46507 papio hamad

ALIGNMENTS

RESULT 1  
Q96FE7 PRELIMINARY; PRT; 263 AA.  
ID Q96FE7  
AC Q96FE7;  
DT 01-DEC-2001 (TRENBLrel. 19, Created)  
DT 01-DEC-2001 (TRENBLrel. 19, Last sequence update)  
DT 01-JUN-2003 (TRENBLrel. 24, Last annotation update)  
DE Hypothetical protein (HGFL(L) protein).  
GN HGFL.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]\_TaxID=9606;  
RP SEQUENCE FROM N.A.  
RC Tissue=Brain;  
RA Strausberg R.;  
RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP SEQUENCE FROM N.A.  
RA Chiang H., Chang M.;  
RL Submitted (JUL-2002) to the EMBL/GenBank/DBJ databases.  
CC -!- SIMILARITY: CONTAINS 1 KRINGLE DOMAIN.  
DR EMBL; BC011049; AAH1049.1; -.  
DR EMBL; AF528080; AAC33763.1; -.  
DR InterPro; IPR000001; Kringle.  
DR Pfam; PF00051; Kringle; 1.  
DR PRINTS; PR00018; Kringle; 1.  
DR ProDom; PD000395; Kringle; 1.  
DR SMART; SM00130; KR; 1.  
DR PROSITE; PS00021; KRINGLE\_1; 1.  
DR PROSITE; PS00070; KRINGLE\_2; 1.  
KW Hypothetical protein; Glycoprotein; Kringle.  
SQ SEQUENCE 263 AA; 28234 MW; 197C3EE888FA242 CRC64;

Query Match 100.0%; Score 1388; DB 4; Length 263;  
Best Local Similarity 100.0%; Pred. NO. 1.6e-119; Indels 0; Gaps 0;  
Matches 262; Conservative 0; Mismatches 0;

```
QY 1 LLAWQAFVLSNMLLAAYSGGCFWNGHLYREDQTSAPGLRCLNWLDAQSLASAPV 60
DB 2 LLAWQAFVLSNMLLAAYSGGCFWNGHLYREDQTSAPGLRCLNWLDAQSLASAPV 61
QY 61 SGAGNHSYCRNPDEDPGRPCWCVSGEAGVPEKPCEDLRCPETTSQALPAFTTEIQEASE 120
DB 62 SGAGNHSYCRNPDEDPGRPCWCVSGEAGVPEKPCEDLRCPETTSQALPAFTTEIQEASE 121
QY 121 GPGADEVOVFAPANALPARSEAAAVQPVIGISQVRMNSKEKKDLGTLGYVLGITMMVII 180
DB 122 GPGADEVOVFAPANALPARSEAAAVQPVIGISQVRMNSKEKKDLGTLGYVLGITMMVII 181
QY 181 IAIGAGIILGYSKRGKDLKEQHDQKVCEREMQRTITPLSAFTNPTCEIVDEKTVVHTS 240
DB 182 IAIGAGIILGYSKRGKDLKEQHDQKVCEREMQRTITPLSAFTNPTCEIVDEKTVVHTS 241
QY 241 QTPVDPQEGSTPLMGQAGTPGA 262
DB 242 QTPVDPQEGSTPLMGQAGTPGA 263

RESULT 2
O00318 PRELIMINARY; PRT; 263 AA.
ID O00318
AC O00318;
DT 01-JUL-1997 (TrEMBLrel. 04, Created)
DT 01-JUL-1997 (TrEMBLrel. 04, Last sequence update)
DE WUGSC:DJ515N1.2 protein.
DE WUGSC:DJ515N1.2 protein.
GN WUGSC:DJ515N1.2
OS Homo sapiens (Human)
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OC NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Du Z., Scheet P., Harper M.;
RT "The sequence of H. sapiens PAC clone RP3-515N1.1";
RL Submitted (MAY-1997) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA Waterston R.;
RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: CONTAINS 1 KRINGLE DOMAIN.
DR EMBL; AC002073; AAB54054.1; -.
DR HSP; P00749; IKDU
DR InterPro; IPR000001; Kringle.
DR Pfam; PF00051; Kringle; 1.
DR PRINTS; PR00018; KRINGLE.
DR PRODOM; PD000395; Kringle; 1.
DR SMART; SM00130; KR; 1.
DR PROSITE; PS00021; KRINGLE_1; 1.
DR PROSITE; PS50070; KRINGLE_2; 1.
KW Glycoprotein; Kringle.
SQ SEQUENCE 263 AA; 28248 MW; 1973EBE8E54A242 CRC64;

Query Match 99.8%; Score 1385; DB 4; Length 263;
Best Local Similarity 99.6%; Pred. No. 3.1e-119;
Matches 261; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 LLAWQAFVLSNMLLAAYSGGCFWNGHLYREDQTSAPGLRCLNWLDAQSLASAPV 60
DB 2 LLAWQAFVLSNMLLAAYSGGCFWNGHLYREDQTSAPGLRCLNWLDAQSLASAPV 61
QY 61 SGAGNHSYCRNPDEDPGRPCWCVSGEAGVPEKPCEDLRCPETTSQALPAFTTEIQEASE 120
DB 62 SGAGNHSYCRNPDEDPGRPCWCVSGEAGVPEKPCEDLRCPETTSQALPAFTTEIQEASE 121
QY 121 GPGADEVOVFAPANALPARSEAAAVQPVIGISQVRMNSKEKKDLGTLGYVLGITMMVII 180
DB 122 GPGADEVOVFAPANALPARSEAAAVQPVIGISQVRMNSKEKKDLGTLGYVLGITMMVII 181
QY 181 IAIGAGIILGYSKRGKDLKEQHDQKVCEREMQRTITPLSAFTNPTCEIVDEKTVVHTS 240
DB 182 IAIGAGIILGYSKRGKDLKEQHDQKVCEREMQRTITPLSAFTNPTCEIVDEKTVVHTS 241
QY 241 QTPVDPQEGSTPLMGQAGTPGA 262
DB 242 QTPVDPQEGSTPLMGQAGTPGA 263

RESULT 3
O00318 PRELIMINARY; PRT; 263 AA.
ID O00318
AC O00318;
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Hypothetical protein FLJ90207.
OS Homo sapiens (Human)
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OC NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Tissue-Mammary gland;
RA Isogai T., Ota T., Nishikawa T., Hayashi K., Otsuki T., Sugiyama T.,
RA Suzuki Y., Nagai K., Sugano S., Ishii S., Kawai-Hio Y., Saito K.,
RA Yamamoto J., Wakamatsu A., Nakamura Y., Kojima S., Nagahari K.,
RA Masuho Y., Ono T., Okano K., Yoshikawa Y., Aotsuka S., Sasaki N.,
RA Hattori A., Okumura K., Iwayanagi T., Ninomiya K.;
RA "NBD0 human cDNA sequencing project.";
RT Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: CONTAINS 1 KRINGLE DOMAIN.
DR EMBL; AK074688; BAC11140.1; -.
DR InterPro; IPR000001; Kringle.
DR Pfam; PF00051; Kringle; 1.
DR PRINTS; PR00018; KRINGLE.
DR PRODOM; PD000395; Kringle; 1.
DR SMART; SM00130; KR; 1.
DR PROSITE; PS00021; KRINGLE_1; 1.
DR PROSITE; PS50070; KRINGLE_2; 1.
KW Hypothetical protein; Glycoprotein; Kringle.
SQ SEQUENCE 263 AA; 28104 MW; 73A9294D5142C63 CRC64;

Query Match 99.0%; Score 1374; DB 4; Length 263;
Best Local Similarity 99.2%; Pred. No. 3.2e-118;
Matches 260; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 LLAWQAFVLSNMLLAAYSGGCFWNGHLYREDQTSAPGLRCLNWLDAQSLASAPV 60
DB 2 LLAWQAFVLSNMLLAAYSGGCFWNGHLYREDQTSAPGLRCLNWLDAQSLASAPV 61
QY 61 SGAGNHSYCRNPDEDPGRPCWCVSGEAGVPEKPCEDLRCPETTSQALPAFTTEIQEASE 120
DB 62 SGAGNHSYCRNPDEDPGRPCWCVSGEAGVPEKPCEDLRCPETTSQALPAFTTEIQEASE 121
QY 121 GPGADEVOVFAPANALPARSEAAAVQPVIGISQVRMNSKEKKDLGTLGYVLGITMMVII 180
DB 122 GPGADEVOVFAPANALPARSEAAAVQPVIGISQVRMNSKEKKDLGTLGYVLGITMMVII 181
QY 181 IAIGAGIILGYSKRGKDLKEQHDQKVCEREMQRTITPLSAFTNPTCEIVDEKTVVHTS 240
DB 182 IAIGAGIILGYSKRGKDLKEQHDQKVCEREMQRTITPLSAFTNPTCEIVDEKTVVHTS 241
QY 241 QTPVDPQEGSTPLMGQAGTPGA 262
DB 242 QTPVDPQEGSTPLMGQAGTPGA 263

RESULT 4
O7TMJ8 PRELIMINARY; PRT; 264 AA.
ID O7TMJ8
AC O7TMJ8;
DT 01-OCT-2003 (TrEMBLrel. 25, Created)
DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
```

```
DB 182 IAIGAGIILGYSKRGKDLKEQHDQKVCEREMQRTITPLSAFTNPTCEIVDEKTVVHTS 241
QY 241 QTPVDPQEGSTPLMGQAGTPGA 262
DB 242 QTPVDPQEGSTPLMGQAGTPGA 263

RESULT 3
O8NCJ9 PRELIMINARY; PRT; 263 AA.
ID O8NCJ9
AC O8NCJ9;
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Hypothetical protein FLJ90207.
OS Homo sapiens (Human)
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OC NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Tissue-Mammary gland;
RA Isogai T., Ota T., Nishikawa T., Hayashi K., Otsuki T., Sugiyama T.,
RA Suzuki Y., Nagai K., Sugano S., Ishii S., Kawai-Hio Y., Saito K.,
RA Yamamoto J., Wakamatsu A., Nakamura Y., Kojima S., Nagahari K.,
RA Masuho Y., Ono T., Okano K., Yoshikawa Y., Aotsuka S., Sasaki N.,
RA Hattori A., Okumura K., Iwayanagi T., Ninomiya K.;
RA "NBD0 human cDNA sequencing project.";
RT Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: CONTAINS 1 KRINGLE DOMAIN.
DR EMBL; AK074688; BAC11140.1; -.
DR InterPro; IPR000001; Kringle.
DR Pfam; PF00051; Kringle; 1.
DR PRINTS; PR00018; KRINGLE.
DR PRODOM; PD000395; Kringle; 1.
DR SMART; SM00130; KR; 1.
DR PROSITE; PS00021; KRINGLE_1; 1.
DR PROSITE; PS50070; KRINGLE_2; 1.
KW Hypothetical protein; Glycoprotein; Kringle.
SQ SEQUENCE 263 AA; 28104 MW; 73A9294D5142C63 CRC64;

Query Match 99.0%; Score 1374; DB 4; Length 263;
Best Local Similarity 99.2%; Pred. No. 3.2e-118;
Matches 260; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 LLAWQAFVLSNMLLAAYSGGCFWNGHLYREDQTSAPGLRCLNWLDAQSLASAPV 60
DB 2 LLAWQAFVLSNMLLAAYSGGCFWNGHLYREDQTSAPGLRCLNWLDAQSLASAPV 61
QY 61 SGAGNHSYCRNPDEDPGRPCWCVSGEAGVPEKPCEDLRCPETTSQALPAFTTEIQEASE 120
DB 62 SGAGNHSYCRNPDEDPGRPCWCVSGEAGVPEKPCEDLRCPETTSQALPAFTTEIQEASE 121
QY 121 GPGADEVOVFAPANALPARSEAAAVQPVIGISQVRMNSKEKKDLGTLGYVLGITMMVII 180
DB 122 GPGADEVOVFAPANALPARSEAAAVQPVIGISQVRMNSKEKKDLGTLGYVLGITMMVII 181
QY 181 IAIGAGIILGYSKRGKDLKEQHDQKVCEREMQRTITPLSAFTNPTCEIVDEKTVVHTS 240
DB 182 IAIGAGIILGYSKRGKDLKEQHDQKVCEREMQRTITPLSAFTNPTCEIVDEKTVVHTS 241
QY 241 QTPVDPQEGSTPLMGQAGTPGA 262
DB 242 QTPVDPQEGSTPLMGQAGTPGA 263

RESULT 4
O7TMJ8 PRELIMINARY; PRT; 264 AA.
ID O7TMJ8
AC O7TMJ8;
DT 01-OCT-2003 (TrEMBLrel. 25, Created)
DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
```

01-OCT-2003 (TRENBLrel. 25, Last annotation update)  
DE Mus musculus (Mouse).  
OS Hypothetical protein.  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=C57BL/6J; TISSUE=Breast tumor;  
RL MEDLINE=22388257; PubMed=12477932;  
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,  
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,  
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,  
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,  
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,  
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,  
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,  
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,  
RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
RA Fahy J., Helton E., Kettman M., Madan A., Rodrigues S., Sanchez A.,  
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,  
RA Krzywinski M.I., Skalska U., Smailus D.E., Schnerch A., Schein J.E.,  
RA Jones S.J., Marra M.A.;  
RT "Generation and initial analysis of more than 15,000 full-length human  
and mouse cDNA sequences.";  
RL Proc. Natl. Acad. Sci. U.S.A. 99:16898-16903(2002).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=C57BL/6J; TISSUE=Breast tumor;  
RA Strausberg R.;  
RL Submitted (AUG-2003) to the EMBL/GenBank/DBJ databases.  
DR EMBL; BC055920; AAH55920.1; -;  
KW Hypothetical protein.  
SQ SEQUENCE 264 AA; 28567 MW; 833EAS78FEB763A4 CRC64;  
  
Query Match 80.4%; Score 1115.5; DB 11; Length 264;  
Best Local Similarity 79.1%; Pred. No. 28-94;  
Matches 208; Conservative 17; Mismatches 37; Indels 1; Gaps 1;  
  
Qy 1 LLAWQAFLVSNMLLAAYGSGCFWNGHLYREDQTSAPGLRCLNWLDAQSLASAPV 60  
Db 2 LLAWHTFLSNMLLAAYGSGCFWNGHLYREDQTSAPGLRCLNWLDAQSLASAPV 61  
Qy 61 SGAGNHSYCRNPDPGRGWCYVSAGVPEKPCEDLKPCTTSA-LPAFTTEIQEAS 119  
Db 62 PSFGNHNYCRNPDPGRGWCYVSAGVPEKPCEDVSCPETTSQAPPPSSAMELEEK 121  
Qy 120 EPGCADEVQVFPANALPARSEAAVQPVIGISQVRMNSKEKKDLGTGLGYVLGITMVI 179  
Db 122 GAPGDKAQQVFPANALPARSEAAVQPVIGISQVRMNSKEKKDLGTGLGYVLGITMVI 181  
Qy 180 IIAIGAGIILGYSYKRGKDLKEQHDQKVCEREMQRTILPLSAFTNPTCEIVDEKTVVHT 239  
Db 182 ILAIGAGIIVGYTYKRGKDLKEQHEKACACEREMQRTILPLSAFTNPTCEIVDEKTVVHT 241  
Qy 240 SQTVPDPQEGSTPLMGAGTPGA 262  
Db 242 NQTPADVQEGSTPLMGAGTPGA 264  
  
RESULT 5  
Q81122 PRELIMINARY; PRT; 264 AA.  
AC Q81122  
DT 01-JUN-2003 (TRENBLrel. 24, Created)  
DT 01-JUN-2003 (TRENBLrel. 24, Last sequence update)  
DT 01-OCT-2003 (TRENBLrel. 25, Last annotation update)  
DE HGFL(L) protein.  
GN HGFL.  
  
Query Match 74.5%; Score 1034; DB 4; Length 234;  
Best Local Similarity 100.0%; Pred. No. 5.4e-87;  
Matches 195; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=BALB/c;  
RL Submitted (JUL-2002) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AF528081; AAC33762.1; -;  
DR InterPro; IPR000001; Kringle.  
DR Pfam; PF00051; Kringle; 1.  
DR PRINTS; PRO0018; KRINGLE.  
DR SMART; SM00130; KR; 1.  
DR PROSITE; PS00021; KRINGLE 1; 1.  
DR PROSITE; PS50070; KRINGLE 2; 1.  
SQ SEQUENCE 264 AA; 28595 MW; 833EAS78FEB50E34 CRC64;  
  
Query Match 80.1%; Score 1111.5; DB 11; Length 264;  
Best Local Similarity 78.7%; Pred. No. 4.6e-94;  
Matches 207; Conservative 18; Mismatches 37; Indels 1; Gaps 1;  
  
Qy 1 LLAWQAFLVSNMLLAAYGSGCFWNGHLYREDQTSAPGLRCLNWLDAQSLASAPV 60  
Db 2 LLAWHTFLSNMLLAAYGSGCFWNGHLYREDQTSAPGLRCLNWLDAQSLASAPV 61  
Qy 61 SGAGNHSYCRNPDPGRGWCYVSAGVPEKPCEDLKPCTTSA-LPAFTTEIQEAS 119  
Db 62 PSFGNHNYCRNPDPGRGWCYVSAGVPEKPCEDVSCPETTSQAPPPSSAMELEEK 121  
Qy 120 EPGCADEVQVFPANALPARSEAAVQPVIGISQVRMNSKEKKDLGTGLGYVLGITMVI 179  
Db 122 GAPGDKAQQVFPANALPARSEAAVQPVIGISQVRMNSKEKKDLGTGLGYVLGITMVI 181  
Qy 180 IIAIGAGIILGYSYKRGKDLKEQHDQKVCEREMQRTILPLSAFTNPTCEIVDEKTVVHT 239  
Db 182 ILAIGAGIIVGYTYKRGKDLKEQHEKACACEREMQRTILPLSAFTNPTCEIVDEKTVVHT 241  
Qy 240 SQTVPDPQEGSTPLMGAGTPGA 262  
Db 242 NQTPADVQEGSTPLMGAGTPGA 264  
  
RESULT 6  
Q86YW2 PRELIMINARY; PRT; 234 AA.  
ID Q86YW2  
AC Q86YW2  
DT 01-JUN-2003 (TRENBLrel. 24, Created)  
DT 01-JUN-2003 (TRENBLrel. 24, Last sequence update)  
DT 01-OCT-2003 (TRENBLrel. 25, Last annotation update)  
DE HGFL(S) protein.  
GN HGFL.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RL Submitted (JUL-2002) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AF528079; AAC33762.1; -;  
DR InterPro; IPR000001; Kringle.  
DR Pfam; PF00051; Kringle; 1.  
DR PRINTS; PRO0018; KRINGLE.  
DR SMART; SM00130; KR; 1.  
DR PROSITE; PS00021; KRINGLE 1; 1.  
DR PROSITE; PS50070; KRINGLE 2; 1.  
SQ SEQUENCE 234 AA; 25320 MW; C78F64DBD1B8DC0D CRC64;  
  
Query Match 74.5%; Score 1034; DB 4; Length 234;  
Best Local Similarity 100.0%; Pred. No. 5.4e-87;  
Matches 195; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
QY 1 LLAWQAFVLSNMLLABAYSGGCFWNGHLYREDQTSAPGLRCLNWLDAQSGLASAPV 60
DB 2 LLAWQAFVLSNMLLABAYSGGCFWNGHLYREDQTSAPGLRCLNWLDAQSGLASAPV 61
QY 61 SGAGNHSCYCRNPDDPRGPGWYVSGAGVPEKPCEDLRCPETTSQALPAFTTIEQASE 120
DB 62 SGAGNHSCYCRNPDDPRGPGWYVSGAGVPEKPCEDLRCPETTSQALPAFTTIEQASE 121
QY 121 GPGDAEVQVFPANALPARSEAAVQPVIGISORVRMNSKEKKDLGTLGYVLGITMWVII 180
DB 122 GPGDAEVQVFPANALPARSEAAVQPVIGISORVRMNSKEKKDLGTLGYVLGITMWVII 181
QY 181 IAIAGAGIILGYSYKR 195
DB 182 IAIAGAGIILGYSYKR 196
RESULT 7
Q81123 PRELIMINARY; PRT; 213 AA.
AC Q81123
DT 01-JUN-2003 (TRENBLrel. 24, Created)
DT 01-JUN-2003 (TRENBLrel. 24, Last sequence update)
DT 01-OCT-2003 (TRENBLrel. 25, Last annotation update)
DE HGFL(S) protein.
GN HGFL
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OC NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BALB/c;
RA Chiang H., Chang M.;
RL Submitted (JUL-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF528078; AAC33761.1; -.
DR InterPro; IPR000001; Kringle.
DR Pfam; PF00051; Kringle; 1.
DR PRINTS; PR00018; KRINGLE.
DR SMART; SM00130; KR; 1.
DR PROSITE; PS00021; KRINGLE 1; 1.
DR PROSITE; PS00070; KRINGLE 2; 1.
SQ SEQUENCE 213 AA; 23011 MW; 4A5E3481025EA597 CRC64;
Query Match 59.2%; Score 821.5; DB 11; Length 213;
Best Local Similarity 78.6%; Pred. No. 1.8e-57;
Matches 154; Conservative 11; Mismatches 30; Indels 1; Gaps 1;
QY 1 LLAWQAFVLSNMLLABAYSGGCFWNGHLYREDQTSAPGLRCLNWLDAQSGLASAPV 60
DB 2 LLAWQAFVLSNMLLABAYSGGCFWNGHLYREDQTSAPGLRCLNWLDAQSGRESLITE 61
QY 61 SGAGNHSCYCRNPDDPRGPGWYVSGAGVPEKPCEDLRCPETTSQA-LPAFTTIEQAS 119
DB 62 PSPGNHNYCRNPDDPRGPGWYVSGAGVPEKPCEDVSCPETTSQAPFPSSAMELEERS 121
QY 120 EGPDAEVQVFPANALPARSEAAVQPVIGISORVRMNSKEKKDLGTLGYVLGITMWVI 179
DB 122 GAPGDKAQVFPANALPARSEAAVQPVIGISOLVRMNSKEKKDLGTLGYVLGITMWVI 181
QY 180 IAIAGAGIILGYSYKR 195
DB 182 IAIAGAGIILGYTYKR 197
```

```
RESULT 8
Q7SXB3 PRELIMINARY; PRT; 263 AA.
AC Q7SXB3
DT 01-OCT-2003 (TRENBLrel. 25, Created)
DT 01-OCT-2003 (TRENBLrel. 25, Last sequence update)
DT 01-OCT-2003 (TRENBLrel. 25, Last annotation update)
DE Hypothetical protein.
```

```
OS Brachydanio rerio (Zebrafish) (Danio rerio).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;
OC Cyprinidae; Danio.
OC NCBI_TaxID=7955;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=AB; TISSUE=Body;
RC MEDLINE=22388257; PubMed=12477932;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Heide F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Lomelianno N.A., Peters G.J., Abramson R.D., Mullaly S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahy J., Helton E., Kettman M., Madan A.C., Rodrigues S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
RA Krzywinski M.I., Skaleka U., Smailus D.E., Schnerch A., Schein J.E.,
RA Jones S.J., Maria M.A.;
RA "Generation and initial analysis of more than 15,000 full-length human
RT and mouse cDNA sequences.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=AB; TISSUE=Body;
RA Strausberg R.;
RL Submitted (AUG-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC055675; AAHS5675.1; -.
KW Hypothetical protein.
SQ SEQUENCE 263 AA; 28777 MW; 8B8EC117EC7C9A58 CRC64;
Query Match 44.0%; Score 611; DB 13; Length 263;
Best Local Similarity 47.9%; Pred. No. 5.5e-48;
Matches 125; Conservative 36; Mismatches 56; Indels 44; Gaps 7;
QY 24 CFWDNGHLYREDQTSAPGLRCLNWLDAQSGLASAPVSGAGNHSCYCRNPDDPR 76
DB 25 CFWDNGHLYREDQTSAPGLRCLNWLDAQSGLASAPVSGAGNHSCYCRNPDDPR 76
QY 77 RGPWCYVSGAGVPEKPCEDLRCPETTSQALPAFTTIEQASEGPGADE----- 126
DB 77 K-EMCVSGSGGCTKEACDIRICQD-----QNATEAPAPESVPTQGLTOR 122
QY 127 -VQVFPANALPARSEAAVQPVIGISORVRMNSKEKKDLGTLGYVLGITMWVIIIGA 185
DB 123 MVETFPANSPFQVEGAAVQPVKGVQVQVSGGPKKKDLGTLGYVLAVFMALLILGG 182
QY 186 GILGYSYKRKDLKEQHQKVCEREMORITLPLSAFTNPTCEIVDEKTVV-----HTSQT 242
DB 183 GITMGVYKGRDLKKQHEORVYEREMHRIPLPLSAFANPTCELVDENTIVITAEPNQT 242
QY 243 PV-DPOEGSTPLMGQAGTGA 262
DB 243 PTQEPVEGADPLMGSAAGTGA 263
RESULT 9
Q8SQ23 PRELIMINARY; PRT; 562 AA.
AC Q8SQ23
DT 01-JUN-2002 (TRENBLrel. 21, Created)
DT 01-JUN-2002 (TRENBLrel. 21, Last sequence update)
DT 01-OCT-2003 (TRENBLrel. 25, Last annotation update)
DE T-plasminogen activator.
OS Sus scrofa (Pig).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
```



```

OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
OX NCBI_TaxID=9823;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Enamel organ;
RA Ding Y., Xue J., Bartlett J.D.;
RT "T-plasminogen activator in tooth tissues.";
RL Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: CONTAINS 2 KRINGLE DOMAINS.
DR EMBL; AF364605; AA00297.1; -.
DR HSP; P00761; IAN1.
DR GO; GO:0005576; C:extracellular; IEA.
DR GO; GO:0004263; F:chymotrypsin activity; IEA.
DR GO; GO:0008233; F:peptidase activity; IEA.
DR GO; GO:0004295; F:trypsin activity; IEA.
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR InterPro; IPR009003; Cys Ser trypsin.
DR InterPro; IPR006209; EGF-like.
DR InterPro; IPR000083; Fibrinctn1.
DR InterPro; IPR006210; IEGF.
DR InterPro; IPR000001; Kringle.
DR InterPro; IPR001254; Peptidase S1.
DR InterPro; IPR001314; Peptidase_S1A.
DR Pfam; PF00008; EGF; 1.
DR Pfam; PF00039; fni; 1.
DR Pfam; PF00051; kringle; 2.
DR Pfam; PF00089; trypsin; 1.
DR PRINTS; PR00722; CHYMOTRYPSIN.
DR PRINTS; PR00018; KRINGLE.
DR ProDom; PD000395; Kringle; 2.
DR SMART; SM00181; EGF; 1.
DR SMART; SM00058; FN1; 1.
DR SMART; SM00130; KR; 2.
DR SMART; SM00020; Tryp Spc; 1.
DR PROSITE; PS00022; EGF_1; 1.
DR PROSITE; PS01186; EGF_2; 1.
DR PROSITE; PS01253; FIBONECTIN_1; 1.
DR PROSITE; PS00021; KRINGLE_1; 2.
DR PROSITE; PS00070; KRINGLE_2; 2.
DR PROSITE; PS02400; TRYPSIN_DOM; 1.
DR PROSITE; PS0134; TRYPSIN_HIS; 1.
DR PROSITE; PS00135; TRYPSIN_SER; 1.
DR EGF-like domain; Glycoprotein; Hydrolase; Kringle; Protease;
KW Serine protease.
SQ SEQUENCE 562 AA; 63668 MW; F9E6B4C77CB101E8 CRC64;

Query Match 12.0%; Score 167; DB 6; Length 562;
Best Local Similarity 36.6%; Pred. No. 1.1e-06;
Matches 41; Conservative 12; Mismatches 33; Indels 26; Gaps 6;

QY 24 CFWDNGHLYREDOTSPAPGLRCLNWLDAQSGLASAPVS-----GAGNHSYCRNPDE 74
DB 127 CYEDQGITVGTWSTAESGAECTNW--NTSGLASMPYNGRRPDAVKLGLGNHNYCENPDK 184

QY 75 DPRGPWCYV-SGAGVPEKPCEDLCFPTTSQALPAFTETIOEASEGPGAD 125
DB 185 DSK-PWCYVF-KAGYSSEFCSTPACSEGS 213

RESULT 10
Q7Z7N2 PRELIMINARY; PRT; 291 AA.
AC Q7Z7N2;
DT 01-OCT-2003 (TrEMBLrel. 25, Created)
DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DE Plasminogen activator, tissue type isoform 2.
GN PLar.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]

```

```

RP SEQUENCE FROM N.A.
RA Rieder M.J., Carrington D.P., da Ponte S.H., Hastings N.C.,
RA Ahearn M.O., Kuldanek S.A., Rajkumar N., Toth E.J., Yi Q.,
RA Nickerson D.A.;
RL Submitted (MAY-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY291060; AAP34246.1; -. 874E38C52F50BF1D CRC64;
SQ SEQUENCE 291 AA; 32191 MW; 874E38C52F50BF1D CRC64;

Query Match 11.1%; Score 154.5; DB 4; Length 291;
Best Local Similarity 39.6%; Pred. No. 6.7e-06;
Matches 36; Conservative 6; Mismatches 36; Indels 13; Gaps 4;

QY 24 CFWDNGHLYREDOTSPAPGLRCLNWLDAQSGLASAPVS-----GAGNHSYCRNPDE 74
DB 127 CYEDQGITVGTWSTAESGAECTNW--NTSGLASMPYNGRRPDAVKLGLGNHNYCENPDK 184

QY 75 DPRGPWCYVSGAGVPEKPCEDLCFPTTS 105
DB 185 DSK-PWCYVF-KAGYSSEFCSTPACSEGS 213

RESULT 11
Q9BU99 PRELIMINARY; PRT; 516 AA.
AC Q9BU99;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DE 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Similar to plasminogen activator, tissue.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Skin;
RA Strausberg R.;
RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1.
CC -1- SIMILARITY: CONTAINS 2 KRINGLE DOMAINS.
DR EMBL; BC002795; AA02795.1; -.
DR HSP; P00750; IASH.
DR GO; GO:0004263; F:chymotrypsin activity; IEA.
DR GO; GO:0008233; F:peptidase activity; IEA.
DR GO; GO:0004295; F:trypsin activity; IEA.
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR InterPro; IPR009003; Cys Ser trypsin.
DR InterPro; IPR006209; EGF-like.
DR InterPro; IPR006210; IEGF.
DR InterPro; IPR000001; Kringle.
DR InterPro; IPR001254; Peptidase S1.
DR InterPro; IPR001314; Peptidase_S1A.
DR Pfam; PF00008; EGF; 1.
DR Pfam; PF00051; kringle; 2.
DR Pfam; PF00089; trypsin; 1.
DR PRINTS; PR00722; CHYMOTRYPSIN.
DR PRINTS; PR00018; KRINGLE.
DR ProDom; PD000395; Kringle; 2.
DR SMART; SM00181; EGF; 1.
DR SMART; SM00130; KR; 2.
DR SMART; SM00020; Tryp Spc; 1.
DR PROSITE; PS00022; EGF_1; 1.
DR PROSITE; PS01186; EGF_2; 1.
DR PROSITE; PS00021; KRINGLE_1; 2.
DR PROSITE; PS00070; KRINGLE_2; 2.
DR PROSITE; PS02400; TRYPSIN_DOM; 1.
DR PROSITE; PS00134; TRYPSIN_HIS; 1.
DR PROSITE; PS00135; TRYPSIN_SER; 1.
DR EGF-like domain; Glycoprotein; Hydrolase; Kringle; Protease;
KW Serine protease.
SQ SEQUENCE 516 AA; 57370 MW; BAB31901FDC96800 CRC64;

Query Match 11.1%; Score 154.5; DB 4; Length 516;

```

```
Best Local Similarity 39.6%; Pred. No. 1.4e-05;
Matches 36; Conservative 6; Mismatches 36; Indels 13; Gaps 4;

QY 24 CFWDNGHLYREDQTSAPGLCLNWLDAQSLASAPVS-----GAGNHSYCRNPDE 74
D 81 CYEQGISYRGTSWTAESGAECTNW--NSSALAKPKYSGRRPDALRGLGNHNYCRNPDR 138
QY 75 DPGPWCYVSGEAGVPEKRPCEDLRCPEPTS 105
D 139 DSK-PWCYVF-KAGKYSSEFCSTACSEGNS 167

RESULT 12
Q86YK8 PRELIMINARY; PRT; 562 AA.
AC Q86YK8;
DT 01-JUN-2003 (TREMBLrel. 24, Created)
DT 01-JUN-2003 (TREMBLrel. 24, Last sequence update)
DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)
DE Tissue plasminogen activator.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Liu Y., Xu L., Zeng Y., He X.;
RT "cDNA of tissue plasminogen activator.";
RL Submitted (JAN-2003) to the EMBL/Genbank/DBSJ databases.
DR EMBL; AY221101; AAC34406.1; -.
DR GO; GO:0005576; C:extracellular; IEA.
DR GO; GO:0004263; F:chymotrypsin activity; IEA.
DR GO; GO:0004295; F:trypsin activity; IEA.
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR InterPro; IPR009003; Cys_Ser_trypsin.
DR InterPro; IPR006209; EGF-like.
DR InterPro; IPR000083; Fibrinctn1.
DR InterPro; IPR000001; Kringle.
DR InterPro; IPR001254; Peptidase_S1.
DR InterPro; IPR001314; Peptidase_S1A.
DR Pfam; PF00039; fn1; 1.
DR Pfam; PF00039; fn1; 1.
DR PRINTS; PR00089; trypsin; 1.
DR PRINTS; PR00722; CHYMOTRYPSIN.
DR ProDom; PD000395; Kringle; 2.
DR SMART; SM00058; FN1; 1.
DR SMART; SM00020; Tryp_Spc; 1.
DR PROSITE; PS00022; EGF_1; 1.
DR PROSITE; PS01186; EGF_2; 1.
DR PROSITE; PS01253; FIBRONECTIN_1; 1.
DR PROSITE; PS00021; KRINGLE_1; 1.
DR PROSITE; PS00021; KRINGLE_2; 1.
DR PROSITE; PS00240; TRYPSIN_DOM; 1.
DR PROSITE; PS00134; TRYPSIN_HIS; 1.
DR PROSITE; PS00135; TRYPSIN_SER; 1.
KW Hypothetical protein; EGF-like domain; Glycoprotein; Hydrolase;
QY SEQUENCE 562 AA; 62902 MW; 837D98392F6EDD1F CRC64;

Query Match 11.1%; Score 154.5; DB 4; Length 562;
Best Local Similarity 39.6%; Pred. No. 1.6e-05;
Matches 36; Conservative 6; Mismatches 36; Indels 13; Gaps 4;

QY 24 CFWDNGHLYREDQTSAPGLCLNWLDAQSLASAPVS-----GAGNHSYCRNPDE 74
D 127 CYEQGISYRGTSWTAESGAECTNW--NSSALAKPKYSGRRPDALRGLGNHNYCRNPDR 184
QY 75 DPGPWCYVSGEAGVPEKRPCEDLRCPEPTS 105
D 185 DSK-PWCYVF-KAGKYSSEFCSTACSEGNS 213
```

```
RESULT 13
Q8VCS4 PRELIMINARY; PRT; 653 AA.
AC Q8VCS4;
DT 01-MAR-2002 (TREMBLrel. 20, Created)
DT 01-MAR-2002 (TREMBLrel. 20, Last sequence update)
DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)
DE Hypothetical protein.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA Strausberg R.;
RL Submitted (DEC-2001) to the EMBL/Genbank/DBSJ databases.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1.
CC -1- SIMILARITY: CONTAINS 1 KRINGLE DOMAIN.
DR EMBL; BC019376; AAH19376.1; -.
DR HSP; P00761; IAN1.
DR GO; GO:0005576; C:extracellular; IEA.
DR GO; GO:0004263; F:chymotrypsin activity; IEA.
DR GO; GO:0008233; F:peptidase activity; IEA.
DR GO; GO:0004295; F:trypsin activity; IEA.
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR InterPro; IPR009003; Cys_Ser_trypsin.
DR InterPro; IPR000742; EGF_2.
DR InterPro; IPR006209; EGF-like.
DR InterPro; IPR000083; Fibrinctn1.
DR InterPro; IPR000582; FN_Type_II.
DR InterPro; IPR006210; IEGF.
DR InterPro; IPR000001; Kringle.
DR InterPro; IPR001254; Peptidase_S1.
DR InterPro; IPR001314; Peptidase_S1A.
DR Pfam; PF00008; EGF; 2.
DR Pfam; PF00039; fn1; 1.
DR Pfam; PF00040; fn2; 1.
DR Pfam; PF00051; kringle; 1.
DR Pfam; PF00089; trypsin; 1.
DR PRINTS; PR00722; CHYMOTRYPSIN.
DR PRINTS; PR00013; FNTYPEII.
DR PRINTS; PR00018; KRINGLE.
DR ProDom; PD000995; FN_Type_II; 1.
DR ProDom; PD000395; Kringle; 1.
DR SMART; SM00181; EGF; 2.
DR SMART; SM00059; FN2; 1.
DR SMART; SM00130; KR; 1.
DR SMART; SM00020; Tryp_Spc; 1.
DR PROSITE; PS00022; EGF_1; 2.
DR PROSITE; PS01186; EGF_2; 1.
DR PROSITE; PS01253; FIBRONECTIN_1; 1.
DR PROSITE; PS00023; FIBRONECTIN_2; 1.
DR PROSITE; PS00021; KRINGLE_1; 1.
DR PROSITE; PS00021; KRINGLE_2; 1.
DR PROSITE; PS00240; TRYPSIN_DOM; 1.
DR PROSITE; PS00134; TRYPSIN_HIS; 1.
DR PROSITE; PS00135; TRYPSIN_SER; 1.
KW Hypothetical protein; EGF-like domain; Glycoprotein; Hydrolase;
QY SEQUENCE 653 AA; 70553 MW; FE18D90174ED6FDD CRC64;

Query Match 10.8%; Score 150; DB 11; Length 653;
Best Local Similarity 33.3%; Pred. No. 5e-05;
Matches 47; Conservative 18; Mismatches 44; Indels 32; Gaps 8;

QY 24 CFWDNGHLYREDQTSAPGLCLNWLDAQSLASAPVS-----LDAQSLASAPVSGAGNHSYCRNPDED 75
D 283 CFLGNGTEYRGVAVSTAAAGLSCLAWNSDLLYQELHVDV-VAAAVLLGLGPHAYCRNPKD 341
QY 76 PRGPWCYVSGEAGVPEKRPCEDLRCPEPTSQALETQIQAEGSGPGAEVQVAFANA 135
D 342 ER-PWCYVVKDNALSWE-----YCRUTACESL-----ARVHQSPP-----EILA---A 380
```

QY 136 LPARSEAAVQPVIGISQVR 156  
 DB 381 LP--ESAPAVPTGGRHKR 399

## RESULT 14

Q8MKB1 PRELIMINARY; PRT; 564 AA.  
 ID Q8MKB1  
 AC Q8MKB1  
 DT 01-OCT-2002 (T-EMBLrel. 22, Created)  
 DT 01-OCT-2002 (T-EMBLrel. 22, Last sequence update)  
 DT 01-OCT-2003 (T-EMBLrel. 25, Last annotation update)  
 DE Tissue-type plasminogen activator.  
 OS Oryctolagus cuniculus (Rabbit).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 CC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.  
 OX NCBI\_TaxID=9986;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Sugiki M., Yoshida E., Anai K., Maruyama M.;  
 RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.  
 CC -I- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1.  
 CC -I- SIMILARITY: CONTAINS 2 KRINGLE DOMAINS.  
 DR EMBL; AY029518; AAK40240.1; .  
 DR GO; GO:0005576; C:extracellular; IEA.  
 DR GO; GO:0004263; F:chymotrypsin activity; IEA.  
 DR GO; GO:0008233; F:peptidase activity; IEA.  
 DR GO; GO:0004295; F:trypsin activity; IEA.  
 DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.  
 DR InterPro; IPR009003; Cys\_Ser\_trypsin.  
 DR InterPro; IPR006209; EGF-like.  
 DR InterPro; IPR000083; Fibnctn1.  
 DR InterPro; IPR006210; IEGF.  
 DR InterPro; IPR000001; Kringle.  
 DR InterPro; IPR001254; Peptidase\_S1.  
 DR InterPro; IPR001314; Peptidase\_S1A.  
 DR Pfam; PF00008; EGF; 1.  
 DR Pfam; PF00039; fn1; 1.  
 DR Pfam; PF00051; kringle; 2.  
 DR Pfam; PF00089; trypsin; 1.  
 DR PRINTS; PR00722; CHYMOTRYPsin.  
 DR PRINTS; PR00018; KRINGLE.  
 DR ProDom; PD000395; Kringle; 2.  
 DR SMART; SM00181; EGF; 1.  
 DR SMART; SM00058; fn1; 1.  
 DR SMART; SM00130; KR; 2.  
 DR SMART; SM00020; Tryp\_SPC; 1.  
 DR PROSITE; PS00022; EGF\_1; 1.  
 DR PROSITE; PS01186; EGF\_2; 1.  
 DR PROSITE; PS01253; FIBRONECTIN\_1; 1.  
 DR PROSITE; PS00021; KRINGLE\_1; 2.  
 DR PROSITE; PS50070; KRINGLE\_2; 2.  
 DR PROSITE; PS02040; TRYPSIN\_DOM; 1.  
 DR PROSITE; PS00134; TRYPSIN\_HIS; 1.  
 DR PROSITE; PS00135; TRYPSIN\_SER; 1.  
 DR EGF-like domain; Glycoprotein; Hydrolase; Kringle; Protease;  
 KW Serine protease.  
 SQ SEQUENCE 564 AA; 62726 MW; 459D8BAC6D4A937C CRC64;

Query Match 10.5%; Score 145.5; DB 6; Length 564;  
 Best Local Similarity 39.5%; Pred. No. 0.00011;  
 Matches 34; Conservative 7; Mismatches 32; Indels 13; Gaps 4;  
 QY 24 CFWDNGHLYREDQTSAPAGRLCLNWLDAQSGLASAPVS-----GAGNHYSYCRNPDE 74  
 DB 128 CYEDRGIGYRGTWSTTESGACQVNW--NSSWLALPKYSGRKPNALRLGLGNHNYCRNPDR 185  
 QY 75 DPRGFWCVSGEAGVPEKPCEDLRC 100  
 DB 186 DTK-PWCYVF-RAGTYSPEFCSTPAC 209

## RESULT 15

Q15146 PRELIMINARY; PRT; 810 AA.  
 ID Q15146  
 AC Q15146  
 DT 01-NOV-1996 (T-EMBLrel. 01, Created)  
 DT 01-NOV-1996 (T-EMBLrel. 01, Last sequence update)  
 DT 01-OCT-2003 (T-EMBLrel. 25, Last annotation update)  
 DE Plasminogen precursor.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 CC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Browne M.J., Chapman C.G., Dodd I., Carey J.E., Lawrence G.M.P.,  
 RA Mitchell D., Robinson J.H.;  
 RT "Expression of recombinant human plasminogen and aglycoplasminogen in  
 RT HeLa cells";  
 RL Fibrinolysis 0:0-0(1991).  
 CC -I- SIMILARITY: CONTAINS 5 KRINGLE DOMAINS.  
 DR EMBL; M74220; AAA36451.1; .  
 DR HSRF; P00747; 2PK4.  
 DR GO; GO:0005509; F:calcium ion binding; IEA.  
 DR GO; GO:0004263; F:chymotrypsin activity; IEA.  
 DR GO; GO:0008233; F:peptidase activity; IEA.  
 DR GO; GO:0003809; F:thrombin activity; IEA.  
 DR GO; GO:0004295; F:trypsin activity; IEA.  
 DR GO; GO:0007596; P:blood coagulation; IEA.  
 DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.  
 DR InterPro; IPR003003; Cys\_Ser\_trypsin.  
 DR InterPro; IPR000001; Kringle.  
 DR InterPro; IPR003014; PAN.  
 DR InterPro; IPR003609; Pan\_app.  
 DR InterPro; IPR001254; Peptidase\_S1.  
 DR InterPro; IPR001314; Peptidase\_S1A.  
 DR InterPro; IPR003966; Peptidase\_S1A\_pr.  
 DR Pfam; PF00051; kringle; 5.  
 DR Pfam; PF00024; PAN; 1.  
 DR Pfam; PF00089; trypsin; 1.  
 DR PRINTS; PR00722; CHYMOTRYPsin.  
 DR PRINTS; PR00018; KRINGLE.  
 DR PRINTS; PR01505; PROTHROMBIN.  
 DR ProDom; PD000395; Kringle; 5.  
 DR SMART; SM00130; KR; 5.  
 DR SMART; SM00473; PAN\_AP; 1.  
 DR SMART; SM00020; Tryp\_SPC; 1.  
 DR PROSITE; PS00021; KRINGLE\_1; 5.  
 DR PROSITE; PS50070; KRINGLE\_2; 5.  
 DR PROSITE; PS50240; TRYPSIN\_DOM; 1.  
 DR PROSITE; PS00134; TRYPSIN\_HIS; 1.  
 DR PROSITE; PS00135; TRYPSIN\_SER; 1.  
 DR Glycoprotein; Hydrolase; Kringle; Protease; Serine protease; Signal.  
 FT SIGNAL 1 19 POTENTIAL.  
 FT CHAIN 20 810 PLASMINOGEN.  
 SQ SEQUENCE 810 AA; 90555 MW; B05C7D4B0D020B3C CRC64;

Query Match 10.2%; Score 142; DB 4; Length 810;  
 Best Local Similarity 37.2%; Pred. No. 0.00036;  
 Matches 32; Conservative 8; Mismatches 34; Indels 12; Gaps 4;  
 QY 24 CFWDNGHLYREDQTSAPAGRLCLNWLDA-----QSGLASAPVSAGNHYSYCRNPDEDPRG 78  
 DB 103 CKTGNGKNGYRGTMSTKNGITCKQKWSSTSPRPSPATHPSEGL-EENYCRNPDPDQ 161  
 QY 79 PWCYVSGEAGVPEK--PCEDLRCE 102  
 DB 162 PWCYTTD-----PEKRYDYCDILECEE 183  
 Search completed: March 17, 2004, 07:06:55  
 Job time : 50.9569 secs

Result No.	Query			DB	ID	Description
	Score	Match	Length			
1	1388	100.0	263	2	Aaw87769	Human t1s
2	1388	100.0	263	2	Aay05219	Kringlet
3	1388	100.0	263	4	Aae00300	Human t1s
4	1388	100.0	263	5	Abr40414	Human sec
5	1385	99.8	263	5	Aau86149	Human PRO
6	1379	99.4	263	3	Aab43437	Human ORF
7	1374	99.0	263	4	Aam93748	Human pol
8	1162.5	83.8	286	2	Aay05220	Kringlet
9	747	53.8	146	5	ABR40487	Human sec
10	747	53.8	146	5	ABR40561	Human sec
11	395	27.7	81	7	ABR42624	Human kri
12	322	23.2	66	4	Aam18900	Peptide #
13	322	23.2	66	4	ABd37905	Peptide #
14	322	23.2	66	4	Aam31314	Peptide #
15	322	23.2	66	4	ABd32159	Protein #
16	322	23.2	66	4	Aam71037	Human bon
17	322	23.2	66	4	Aam58537	Human liv
18	322	23.2	66	4	ABG52752	Human bra
19	322	23.2	66	5	ABG40828	Human pep
20	301	21.7	55	2	Aay12615	Human 5'
21	300	21.6	55	2	Aay12397	Human 5'
22	196	14.1	39	2	Aaw72641	Nervous 9
23	192	13.8	39	2	Aam72640	Nervous 9
24	160.5	11.6	527	2	Aar20220	t-PA anal
25	160.5	11.6	527	2	Aar20219	t-PA anal

XX WO9854199-A1.  
 XX 03-DEC-1998.  
 XX 27-MAY-1998; 98WO-US010728.  
 XX 28-MAY-1997; 97US-0048000P.  
 XX (HUMA-) HUMAN GENOME SCI INC.  
 XX Ebner R, Moore PA, Ruben SM;  
 XX WPI; 1999-070207/06.  
 XX N-PSDB; AAV99636.  
 XX New tissue plasminogen activator-like protease - useful in the diagnosis  
 XX and treatment of circulatory system-related disorders.  
 XX Claim 1; Page 56-57; 76pp; English.  
 XX This is the amino acid sequence of tissue plasminogen activator-like  
 XX protease (t-PALP), a novel member of the serine protease family that  
 XX shares sequence homology to human tissue plasminogen activator (see  
 XX AAW87770). The t-PALP sequence was deduced from a cDNA clone (see  
 XX AAV99636) derived from activated monocytes. The 2.5 kb t-PALP message has  
 XX also been detected in heart, brain, lung, placenta, liver, skeletal  
 XX muscle, kidney, pancreas, spleen, thymus, prostate, testis, ovary, small  
 XX intestine, colon and peripheral blood leukocytes. Isolated nucleic acids  
 XX encoding amino acids -21 to 242, -20 to 242, 1-242, 4-63 (kringle domain)  
 XX and 64-242 (protease domain) of t-PALP, or encoding epitope-bearing  
 XX portions of t-PALP, are also claimed, as are recombinant vectors, host  
 XX cells, and methods for producing t-PALP polypeptides. t-PALP may be used  
 XX to detect and treat disorders related to the circulatory system, and to  
 XX identify agonists and antagonists of t-PALP activity. The homology  
 XX between t-PALP and tPA indicates that t-PALP may be involved in the  
 XX regulation of normal and abnormal clotting in e.g. stroke, deep-vein  
 XX thrombosis, peripheral arterial occlusion, pulmonary embolism and  
 XX myocardiothrombosis  
 XX Sequence 263 AA;  
 XX  
 XX Query Match 100.0%; Score 1388; DB 2; Length 263;  
 XX Best Local Similarity 100.0%; Pred. No. 6.9e-119;  
 XX Matches 262; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 LLAWQAFVLSNMLAEAYGSGGCFWNGHLYREDQTSAPGLRCLNWLDAQSLAPV 60  
 DB 2 LLAWQAFVLSNMLAEAYGSGGCFWNGHLYREDQTSAPGLRCLNWLDAQSLAPV 61  
 QY 61 SGAGNHSYCRNPDEDPRGPWCYVSGAGVPEKPCEDLRCPTTSQALPAFTTEIQEASE 120  
 DB 62 SGAGNHSYCRNPDEDPRGPWCYVSGAGVPEKPCEDLRCPTTSQALPAFTTEIQEASE 121  
 QY 121 GPGADEVQVFPANALPARSEAAVQVIGISQVRMNSKEKDLGTLGYVLGTTMVI 180  
 DB 122 GPGADEVQVFPANALPARSEAAVQVIGISQVRMNSKEKDLGTLGYVLGTTMVI 181  
 QY 181 IAIGAGIILGYSGKXKDLKEHQDKVCEREMQRTITLPLSAFTNPTCEIVDEKTVVHTS 240  
 DB 182 IAIGAGIILGYSGKXKDLKEHQDKVCEREMQRTITLPLSAFTNPTCEIVDEKTVVHTS 241  
 QY 241 QTPVDPQEGSTPLMGAGTPGA 262  
 DB 242 QTPVDPQEGSTPLMGAGTPGA 263  
 RESULT 2  
 AAY05219  
 ID AAY05219 standard; protein; 263 AA.  
 XX  
 XX AC AAY05219;  
 XX

DT 17-JUN-1999 (first entry)  
 XX Kringle1 protein sequence.  
 XX Kringle1; cancer; inflammation; autoimmunity; rheumatoid arthritis; AIDS;  
 XX CNS inflammation; cerebellar degeneration; Alzheimer's disease; asthma;  
 XX Parkinson's disease; multiple sclerosis; amyotrophic lateral sclerosis;  
 XX neurological abnormality; ischaemia reperfusion injury; ischaemic anaemia;  
 XX cardiovascular disease; kidney disease; liver disease; aplastic anaemia;  
 XX myocardial infarction; hypotension; hypertension; allergy; infection;  
 XX myelodysplastic syndrome; haematologic abnormality; diagnosis; therapy;  
 XX male pattern baldness.  
 XX Homo sapiens.  
 XX WO9911788-A1.  
 XX 11-MAR-1999.  
 XX 02-SEP-1998; 98WO-US018270.  
 XX 02-SEP-1997; 97US-0056032P.  
 XX 01-SEP-1998; 98US-00144889.  
 XX (SMIK) SMITHKLINE BEECHAM CORP.  
 XX Albone EF, Kikly KK;  
 XX WPI; 1999-214707/18.  
 XX N-PSDB; AAX28354.  
 XX New kringle1 polypeptides and polynucleotides.  
 XX Claim 1; Page 31-32; 42pp; English.  
 XX This sequence is a Kringle1 polypeptide of the invention. The kringle1  
 XX polypeptides (I) are used to screen for agonists and antagonists.  
 XX Agonists are used to treat subjects in need of enhanced activity or  
 XX expression of (I). Antagonists are used to treat subjects having need to  
 XX inhibit the activity or expression of (I). The methods can be used to  
 XX treat conditions such as cancer, inflammation, autoimmunity, allergy,  
 XX asthma, rheumatoid arthritis, CNS inflammation, cerebellar degeneration,  
 XX Alzheimer's disease, Parkinson's disease, multiple sclerosis,  
 XX amyotrophic lateral sclerosis, head injury damage and other neurological  
 XX abnormalities, ischaemia reperfusion injury, cardiovascular disease,  
 XX kidney disease, liver disease, ischaemic injury, myocardial infarction,  
 XX hypotension, hypertension, aplastic anaemia, male pattern baldness, and  
 XX haematologic abnormalities, aplastic anaemia, viral infections. The kringle1  
 XX bacterial, fungal, protozoan and viral antibodies. Determining the  
 XX polypeptides may also be used to generate antibodies. Determining the  
 XX presence or absence of mutations in, and analysing for the presence or  
 XX absence of expression of, kringle1 polynucleotides can be used to  
 XX diagnose a disease or susceptibility to a disease related to expression  
 XX or activity of kringle1 proteins. The polynucleotides may also be used  
 XX for chromosome identification, and mapping  
 XX Sequence 263 AA;  
 XX  
 XX Query Match 100.0%; Score 1388; DB 2; Length 263;  
 XX Best Local Similarity 100.0%; Pred. No. 6.9e-119;  
 XX Matches 262; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 LLAWQAFVLSNMLAEAYGSGGCFWNGHLYREDQTSAPGLRCLNWLDAQSLAPV 60  
 DB 2 LLAWQAFVLSNMLAEAYGSGGCFWNGHLYREDQTSAPGLRCLNWLDAQSLAPV 61  
 QY 61 SGAGNHSYCRNPDEDPRGPWCYVSGAGVPEKPCEDLRCPTTSQALPAFTTEIQEASE 120  
 DB 62 SGAGNHSYCRNPDEDPRGPWCYVSGAGVPEKPCEDLRCPTTSQALPAFTTEIQEASE 121  
 QY 121 GPGADEVQVFPANALPARSEAAVQVIGISQVRMNSKEKDLGTLGYVLGTTMVI 180  
 DB 122 GPGADEVQVFPANALPARSEAAVQVIGISQVRMNSKEKDLGTLGYVLGTTMVI 181

QY 181 IAIGAGIILGYXKXKDLKEQHDQKVCEREMORITLPLSAFTNPTCEIVDEKTVVHTS 240  
 |||||  
 Db 182 IAIGAGIILGYXKXKDLKEQHDQKVCEREMORITLPLSAFTNPTCEIVDEKTVVHTS 241  
 |||||  
 QY 241 QTPVDPQEGSTPLMGAGTPGA 262  
 |||||  
 Db 242 QTPVDPQEGSTPLMGAGTPGA 263  
 |||||

## RESULT 3

AAE00300

ID AAE00300 standard; protein; 263 AA.

XX AC AAE00300;

XX DT 13-JUN-2001 (first entry)

XX DE Human tissue-plasminogen activator-like protease (t-PALP).

XX KW Human; tissue-plasminogen activator-like protease; t-PALP; therapy;

XX KW vascular disease; stroke; deep vein thrombosis; keloid; asthma;

XX KW arterial occlusion; blood coagulation disorder; cerebroprotective;

XX KW autoimmune system disorder; human immunodeficiency syndrome; cytostatic;

XX KW rheumatoid arthritis; graft-versus-host disease; thyroiditis; cardiac;

XX KW insulin dependent diabetes; inflammatory eye disease; allergy; antiviral;

XX KW cardiovascular disease; heart disease; arrhythmia; myocardial ischaemia;

XX KW hyperproliferative disorder; hypertrophic scar; neurological disease;

XX KW Creutzfeldt-Jakob syndrome; neurodegenerative disorder; antiinflammatory;

XX KW Alzheimer's disease; Parkinson's disease; immunosuppressive; antifungal;

XX KW infectious disease; drug screening; gene therapy; neuroprotective;

XX KW cancer; ophthalmological; antibacterial; vulnertary.

XX KW Homo sapiens.

XX OS

XX Key

XX Binding-site

XX Location/Qualifiers

XX 1..165

XX /note= "Binds to FLAG polypeptide to form t-PALP-FLAG

XX fusion protein"

XX Peptide

XX 1..21

XX /label= Signal peptide

XX Domain

XX 4..63

XX /label= Kringle\_domain

XX Region

XX 12..21

XX /note= "Conserved region"

XX Protein

XX 22..263

XX /note= "Human mature tissue-plasminogen activator-like

XX protease (t-PALP); Binds to FLAG polypeptide to form t-

XX PALP-FLAG fusion protein"

XX Region

XX 22..38

XX /note= "Conserved region"

XX Region

XX 22..31

XX /note= "Epitope-bearing portion"

XX Region

XX 35..44

XX /note= "Epitope-bearing portion"

XX Region

XX 39..49

XX /note= "Conserved region"

XX Region

XX 50..62

XX /note= "Conserved region"

XX Region

XX 63..84

XX /note= "Conserved region"

XX Domain

XX 64..242

XX /label= Protease\_domain

XX Region

XX 71..81

XX /note= "Epitope-bearing portion"

XX Region

XX 85..97

XX /note= "Conserved region"

XX Region

XX 91..107

XX /note= "Epitope-bearing portion"

XX Region

XX 100..118

XX /note= "Conserved region"

XX Region

XX 119..128

XX /note= "Epitope-bearing portion"

XX FT

FT Region 119..127  
 /note= "Conserved region"  
 FT Region 128..143  
 /note= "Conserved region"  
 FT Region 138..147  
 /note= "Epitope-bearing portion"  
 FT Region 146..163  
 /note= "Conserved region"  
 FT Region 155..167  
 /note= "Epitope-bearing portion"  
 FT Region 164..180  
 /note= "Conserved region"  
 FT Region 186..200  
 /note= "Conserved region"  
 FT Region 193..203  
 /note= "Epitope-bearing portion"  
 FT Region 201..220  
 /note= "Conserved region"  
 FT Region 206..215  
 /note= "Epitope-bearing portion"  
 FT Region 221..236  
 /note= "Conserved region"  
 FT Region 227..237  
 /note= "Epitope-bearing portion"  
 FT Region 237..248  
 /note= "Conserved region"  
 FT Region 243..252  
 /note= "Epitope-bearing portion"  
 FT Region 249..263  
 /note= "Conserved region"

WO200125252-A1.

12-APR-2001.

03-OCT-2000; 2000WO-US027239.

04-OCT-1999; 99US-00411977.

(HUMA-) HUMAN GENOME SCI INC.

Moore PA, Ruben SM, Ebner R;

WPI; 2001-235402/24.

N-PSDB; AAD03460.

New (gene encoding and antibody immunospecific for a) tissue-plasminogen activator-like protease, useful for the diagnosis and treatment of (cardio)vascular diseases, hyperproliferative disorders, immune system disorders and cancers.

Claim 17; Fig 1; 323pp; English.

The present amino acid sequence is HMSIB42 clone human tissue-plasminogen activator-like protease (t-PALP). The t-PALP sequence and their (ant)agonists are useful for the diagnosis and treatment of vascular diseases e.g. stroke, deep vein thrombosis and arterial occlusion, blood coagulation disorders, (auto)immune system disorders e.g. human immunodeficiency syndrome, rheumatoid arthritis, graft-versus-host disease, thyroiditis, insulin dependent diabetes and inflammatory eye disease, allergic reactions e.g. asthma, cardiovascular diseases e.g. heart disease, arrhythmia and myocardial ischaemia, hyperproliferative disorders, cancers, hypertrophic scars and keloids, neurological diseases e.g. Creutzfeldt-Jakob syndrome, neurodegenerative disorders e.g. Alzheimer's disease and Parkinson's disease and infectious disease e.g. viral, bacterial and fungal infections. The t-PALP sequences are also useful for drug screening. The t-PALP nucleotides are useful as chromosome markers and are involved in gene therapy

SQ Sequence 263 AA;

Query Match 100.0%; Score 1388; DB 4; Length 263;  
 Best Local Similarity 100.0%; Pred. No. 6.9e-119;

Matches 262; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LLAWQAFVLSNMLLAAYGSGCFWNGHLYREDQTSAPGLRCLNWLDAQSLASAPV 60

Db 2 LLAWQAFVLSNMLLAAYGSGCFWNGHLYREDQTSAPGLRCLNWLDAQSLASAPV 61

QY 61 SGAGNHSYCRNPDEPRGPWCYVSGEAGVPEKPCEDLRCPETTSQALPAFTTIOEASE 120

Db 62 SGAGNHSYCRNPDEPRGPWCYVSGEAGVPEKPCEDLRCPETTSQALPAFTTIOEASE 121

QY 121 GFGADEVQVFPANALPARSEAAAVQVIGISQVRMNSKEKKDLGTLGYVLGTTMMVII 180

Db 122 GFGADEVQVFPANALPARSEAAAVQVIGISQVRMNSKEKKDLGTLGYVLGTTMMVII 181

QY 181 IAGAGIILGYSGKDKLKEQHDQKVCEREMQRTILPLSAFTNPTCEIVDEKTVVHTS 240

Db 182 IAGAGIILGYSGKDKLKEQHDQKVCEREMQRTILPLSAFTNPTCEIVDEKTVVHTS 241

QY 241 QTPVDPQEGSTPLMGQAGTPGA 262

Db 242 QTPVDPQEGSTPLMGQAGTPGA 263

RESULT 4

ABR40414

ID ABR40414 standard; protein; 263 AA.

AC ABR40414;

XX

XX

DT 13-JUN-2003 (first entry)

XX

XX

DE Human secreted protein #SEQ ID 164.

XX

XX

KW Human; secreted protein; anti-HIV; nootropic; neuroprotective;

KW antianginal; immunosuppressive; immunomodulator; cytostatic; cardiant;

KW hepatotropic; antiinflammatory; antiallergic; antidiabetic;

KW gastrointestinal; antinfertility; nephrotropic; virucide; hypotensive;

KW vasotropic; dermatological; osteopathic; antiarthritic; antiparkinsonian;

KW antiasthmatic; antipsoriatic; cerebroprotective; antibacterial;

KW fungicide; antiparasitic; protein therapy; gene therapy; cancer;

KW hyperproliferative disorder; leukaemia; autoimmune disorder;

KW immunodeficiency; HIV; infertility; uterine disorders; endometriosis;

KW amenorrhea; ocular disorder; neurological disorder; wound healing;

KW Huntington's disease; gastrointestinal disorder; inflammatory disease;

KW Crohn's disease; kidney disorder; renal failure; diabetes; obesity;

KW cirrhosis; hepatitis; cardiovascular disorder; angina; osteoarthritis;

KW rheumatoid arthritis; psoriasis; multiple sclerosis; asthma;

KW respiratory disorder; infectious disease; chromosome identification;

KW food additive; nutrition.

XX

OS Homo sapiens.

XX

XX WO200268628-A1.

XX

XX

PD 06-SEP-2002.

XX

XX

PP 21-FEB-2002; 2002WO-US005301.

XX

XX

PR 23-FEB-2001; 2001US-0270625P.

PR

PR 12-JUL-2001; 2001US-0304417P.

XX

XX

PA (HUMA-) HUMAN GENOME SCI INC.

XX

XX Rosen CA, Komatsoulis G, Baker KP, Fiscella M, Moore PA, Wei P;

PI Duan DR, Shi Y, Gupta R;

PI

XX WPI; 2002-750417/81.

DR

DR N-PSDB; ABR282469.

XX

XX New human secreted proteins and nucleic acids, useful for preventing,

PT treating, or ameliorating a medical condition, e.g. cancer, autoimmune

PT disorder, HIV, Parkinson's disease, Alzheimer's disease, diabetes,

PT obesity or cirrhosis.

PT

XX Claim 11; Page 755; 873pp; English.

XX The invention relates to novel human secreted proteins and the genes encoding them. Genes and proteins of the invention may be useful for preventing, treating or ameliorating medical conditions e.g. by protein or gene therapy. These conditions include cancer and hyperproliferative disorders, immune cell proliferative disorders (e.g. leukaemia), autoimmune disorders, immunodeficiencies (e.g. HIV infection), infertility, placental and uterine disorders (e.g. endometriosis), amenorrhea, ocular disorders, neurological disorders (e.g. Huntington's disease), wound healing, gastrointestinal system disorders, particularly inflammatory diseases (e.g. Crohn's disease), kidney disorders (e.g. renal failure), diabetes, obesity, cirrhosis, hepatitis, cardiovascular disorders (e.g. angina), rheumatoid arthritis, osteoarthritis, psoriasis, multiple sclerosis, respiratory disorders (e.g. asthma), or infectious diseases caused by bacterial, parasitic, viral or fungal agents. The nucleic acids are also useful for chromosome identification, radiation hybrid mapping or long-range restriction mapping. The polypeptide, or polynucleotide, agonist or antagonist may also be used as a food additive or preservative to increase or decrease storage capabilities, fat content or other nutritional components. The sequences given in records ABR40409-ABR40590 and ABR282464-ABR282611 represent human secreted proteins and the genes encoding them

XX

SQ Sequence 263 AA;

Query Match 100.0%; Score 1388; DB 5; Length 263;

Best Local Similarity 100.0%; Pred. No. 6.9e-119;

Matches 262; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LLAWQAFVLSNMLLAAYGSGCFWNGHLYREDQTSAPGLRCLNWLDAQSLASAPV 60

Db 2 LLAWQAFVLSNMLLAAYGSGCFWNGHLYREDQTSAPGLRCLNWLDAQSLASAPV 61

QY 61 SGAGNHSYCRNPDEPRGPWCYVSGEAGVPEKPCEDLRCPETTSQALPAFTTIOEASE 120

Db 62 SGAGNHSYCRNPDEPRGPWCYVSGEAGVPEKPCEDLRCPETTSQALPAFTTIOEASE 121

QY 121 GFGADEVQVFPANALPARSEAAAVQVIGISQVRMNSKEKKDLGTLGYVLGTTMMVII 180

Db 122 GFGADEVQVFPANALPARSEAAAVQVIGISQVRMNSKEKKDLGTLGYVLGTTMMVII 181

QY 181 IAGAGIILGYSGKDKLKEQHDQKVCEREMQRTILPLSAFTNPTCEIVDEKTVVHTS 240

Db 182 IAGAGIILGYSGKDKLKEQHDQKVCEREMQRTILPLSAFTNPTCEIVDEKTVVHTS 241

QY 241 QTPVDPQEGSTPLMGQAGTPGA 262

Db 242 QTPVDPQEGSTPLMGQAGTPGA 263

RESULT 5

AAU86149

ID AAU86149 standard; protein; 263 AA.

XX

XX

AC AAU86149;

XX

XX

DT 15-JUL-2002 (first entry)

XX

XX

DE Human PRO264 polypeptide.

XX

XX Human; PRO; benign tumour; malignant tumour; lymphoid malignancy;

KW leukaemia; neuronal disorder; stromal disorder; blastocoele disorder;

KW inflammatory disorder; immune disorder; angiogenic disorder; cytostatic;

KW neuroprotective.

XX

XX Homo sapiens.

XX

XX WO200153486-A1.

XX

XX 26-JUL-2001.

XX

PF 11-FEB-2000; 2000WO-US003565.  
 XX 08-MAR-1999; 99WO-US005028.  
 PR 11-MAR-1999; 99US-0123972P.  
 PR 11-MAY-1999; 99US-0133459P.  
 PR 02-JUN-1999; 99WO-US012252.  
 PR 22-JUN-1999; 99US-0140650P.  
 PR 22-JUN-1999; 99US-0140653P.  
 PR 20-JUL-1999; 99US-0144759P.  
 PR 26-JUL-1999; 99US-0145699P.  
 PR 28-JUL-1999; 99US-0146222P.  
 PR 17-AUG-1999; 99US-0149399P.  
 PR 31-AUG-1999; 99US-0151689P.  
 PR 01-SEP-1999; 99WO-US020111.  
 PR 15-SEP-1999; 99WO-US021090.  
 PR 30-NOV-1999; 99WO-US028313.  
 PR 01-DEC-1999; 99WO-US028301.  
 PR 01-DEC-1999; 99WO-US028634.  
 PR 05-JAN-2000; 2000WO-US000219.  
 XX (GETH ) GENENTECH INC.  
 XX Ashkenazi AJ, Goddard A, Godowski FJ, Gurney AL, Hillan KJ;  
 PI Marsters SA, Pan J, Pitti RM, Roy MA, Smith V, Stone DM;  
 PI Watanabe CX, Wood WI;  
 XX WPI; 2002-205557/26.  
 DR N-PSDB; ABK40275.  
 XX  
 PT Thirty five nucleic acids encoding PRO polypeptides, useful for treating  
 PT benign or malignant tumors, leukemias and lymphoid malignancies,  
 PT inflammatory, angiogenic and immunologic disorders.  
 XX Claim 61; Fig 44; 302pp; English.  
 XX  
 CC The present invention relates to the isolation of novel human PRO  
 CC polypeptides and the polynucleotide sequences encoding them. The PRO  
 CC polypeptides, agonists, antagonists or anti-PRO antibodies are useful for  
 CC treating benign or malignant tumours (e.g. renal, kidney, bladder,  
 CC breast, etc), leukemias and lymphoid malignancies, other disorders such  
 CC as neuronal, glial, astrocytic, hypothalamic, glandular, macrophagal,  
 CC stromal and blastocoeic disorders, inflammatory, immune and angiogenic  
 CC disorders. The polynucleotide sequences are also useful in gene therapy.  
 CC AAU86128-AAU86162 represent the human PRO polypeptides of the invention  
 XX  
 SQ Sequence 263 AA;  
 Query Match 99.8%; Score 1385; DB 5; Length 263;  
 Best Local Similarity 99.8%; Pred. NO. 1.3e-118;  
 Matches 261; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 LLAWVQAFVSNMLLAAYSGGCFWNGHLYREDQTSAPGLRCLNWLDAQSGLASAPV 60  
 Db 2 LLAWVQAFVSNMLLAAYSGGCFWNGHLYREDQTSAPGLRCLNWLDAQSGLASAPV 61  
 QY 61 SGAGNSHCNPNDEPRGWCYVSGAGVPEKPCEDLRCPETTSQALPAFTTEQBASE 120  
 Db 62 SGAGNSHCNPNDEPRGWCYVSGAGVPEKPCEDLRCPETTSQALPAFTTEQBASE 121  
 QY 121 GPQADEVQVFPANALPARSEAAVQPVIGISORVMNSKEKDLGTGLGVLTGMVII 180  
 Db 122 GPQADEVQVFPANALPARSEAAVQPVIGISORVMNSKEKDLGTGLGVLTGMVII 181  
 QY 181 IATGAGIILGYSYKRGKDLKEQHDQKVCEREMQRIITPLSAFTNPCEIVDEKTVVHTS 240  
 Db 182 IATGAGIILGYSYKRGKDLKEQHDQKVCEREMQRIITPLSAFTNPCEIVDEKTVVHTS 241  
 QY 241 QTPVDPQEGSTPLMGAGTGA 262  
 Db 242 QTPVDPQEGSTPLMGAGTGA 263

RESULT 6

AAB43237  
 ID AAB43237 standard; protein; 263 AA.  
 XX  
 AC AAB43237;  
 XX  
 DT 08-FEB-2001 (first entry)  
 XX  
 DE Human ORFX ORF3001 polypeptide sequence SEQ ID NO:6002.  
 XX  
 KW Human; open reading frame; ORFX; detection; cytostatic; hepatotropic;  
 KW vulnary; antipsoriatic; antiparkinsonian; nootropic; neuroprotective;  
 KW anticonvulsant; osteopathic; antiarthritic; immunosuppressant; cardiant;  
 KW immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;  
 KW hypotensive; dermatological; immunosuppressive; antiinflammatory;  
 KW antiviral; antibacterial; antifungal; antineumatic; antichyroid;  
 KW antianemic; gene therapy; cancer; proliferative disorder; hypercension;  
 KW neurodegenerative disorder; osteoarthritis; graft vs host disease;  
 KW cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;  
 KW cholesterol ester storage; systemic lupus erythematosus; infection;  
 KW severe combined immunodeficiency; malaria; autoimmune disorder; asthma;  
 KW allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;  
 KW bone damage; cartilage damage; antinflammatory disease; coagulation;  
 KW thrombosis; contraceptive.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200058473-A2.  
 XX  
 PD 05-OCT-2000.  
 XX  
 PF 31-MAR-2000; 2000WO-US008621.  
 XX  
 PR 31-MAR-1999; 99US-0127607P.  
 PR 02-APR-1999; 99US-0127636P.  
 PR 05-APR-1999; 99US-0127728P.  
 PR 30-MAR-2000; 2000US-00540763.  
 XX  
 XX (CURA-) CURAGEN CORP.  
 XX  
 XX Shinkets RA, Leach M;  
 WPI; 2000-602362/57.  
 N-PSDB; AAC77446.  
 Novel nucleic acids and peptides derived from open reading frame X,  
 useful for treating e.g. cancers, proliferative disorders,  
 neurodegenerative disorders and cardiovascular disease.  
 Claim 11; Page 5181-5182; 5507pp; English.  
 AAC7446 to AAC77606 encode the proteins given in AAB40237 to AAB43387,  
 which represent the human ORFX open reading frames 1 to 3161. The ORFX  
 sequences have activities such as: cytostatic; hepatotropic; vulnary;  
 antipsoriatic; antiparkinsonian; nootropic; neuroprotective; osteopathic;  
 anticonvulsant; antiarthritic; immunosuppressant; immunostimulant;  
 cardiant; thrombolytic; coagulant; vasotropic; antidiabetic; hypotensive;  
 dermatological; immunosuppressive; antiinflammatory; antibacterial;  
 antiviral; antifungal; antineumatic; antichyroid; and antianemic. The  
 sequences can be used for determining the presence of or predisposition  
 to, or preventing or treating pathological conditions associated with an  
 ORFX-associated disorder. The nucleic acids can be used to express ORFX  
 proteins in gene therapy vectors. The proteins and nucleic acids may be  
 used to treat cancers, proliferative disorders, neurodegenerative  
 disorders, osteoarthritis, graft vs host disease, cardiovascular disease,  
 diabetes mellitus, hypercension, hypothyroidism, cholesterol ester  
 storage, systemic lupus erythematosus, severe combined immunodeficiency  
 (SCID), AIDS, viral, bacterial or fungal infection, malaria, autoimmune  
 disorders, asthma, allergies, aplastic anaemia, burns, wounds, bone and  
 cartilage damage, nocturnal haemoglobinuria, antinflammatory disease; to  
 enhance coagulation; to inhibit thrombosis; and as a contraceptive  
 Sequence 263 AA;



Query Match 99.4%; Score 1379; DB 3; Length 263;  
Best Local Similarity 99.2%; Pred. No. 4.6e-118;  
Matches 260; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 LLAWQAFVLSNMLLAAYGSGGCFWNGHLYREDQTSAPGLRCLNWLDAQSLASAPV 60  
DB 2 LLAWQAFVLSNMLLAAYGSGGCFWNGHLYREDQTSAPGLRCLNWLDAQSLASAPV 61

QY 61 SGAGNHSYCRNPDEDPGPGWCYVSGEAGVPEKPCEDLRCPETTSQALPAFTTEIQEASE 120  
DB 62 SGAGNHSYCRNPDEDPGPGWCYVSGEAGVPEKPCEDLRCPETTSQALPAFTTEIQEASE 121

QY 121 GPGADEVQVFPANALPARSEAAAVQPVIGISQVRMNSKEKKDLGTLGYVLGTTMMVII 180  
DB 122 GPGADEVQVFPANALPARSEAAAVQPVIGISQVRMNSKEKKDLGTLGYVLGTTMMVII 181

QY 181 IATGAGIILGYSGKGLKEQHDQKVCEREMQRTITPLSAFTNPTCEIVDEKTVVHTS 240  
DB 182 IATGAGIILGYSGKGLKEQHDQKVCEREMQRTITPLSAFTNPTCEIVDEKTVVHTS 241

QY 241 QTFVDPQEGSTPLMGQAGTPGA 262  
DB 242 QTFVDPQEGSTPLMGQAGTPGA 263

RESULT 7  
AAM93748  
ID AAM93748 standard; protein; 263 AA.  
AC AAM93748;  
XX  
DT 06-NOV-2001 (first entry)  
XX  
DE Human polypeptide, SEQ ID NO: 3727.  
XX  
KW Human; full length cDNA; cDNA synthesis; oligo-capping.  
XX  
OS Homo sapiens.  
XX  
PN EPI130094-A2.  
XX  
XX  
PD 05-SEP-2001.  
XX  
XX  
PF 07-JUL-2000; 2000EP-00114089.  
XX  
PR 08-JUL-1999; 99JP-00194486.  
PR 11-JAN-2000; 2000JP-00118774.  
PR 02-MAY-2000; 2000JP-00183765.  
XX  
XX (HELI-) HELIX RES INST.  
XX  
XX Ota T, Nishikawa T, Isogai T, Hayashi K, Ishii S, Kawai Y;  
PI Wakamatsu A, Sugiyama T, Nagai K, Kojima S, Otsuki T, Koga H;  
XX  
XX WPI; 2001-524255/58.  
DR N-PSDB; AAK94700.  
XX  
XX 830 Primers useful for synthesizing full length cDNA clones and their use  
PT in genetic manipulation.  
XX  
XX Claim 8; SEQ ID NO 3727; 1380pp + Sequence Listing; English.  
PS  
XX  
XX The invention relates to primers for synthesizing full length cDNA  
CC clones. 830 cDNA molecules encoding a human protein have been isolated  
CC and nucleotide sequences of 5'- and 3'-ends of the cDNA molecules have  
CC been determined. Primers for synthesizing the full length cDNA are useful  
CC for clarifying the function of the protein encoded by the cDNA. The full  
CC length clones were obtained by construction of full length enriched cDNA  
CC libraries that were synthesised by the oligo-capping method. The primers  
CC enable the production of the full length cDNA easily without any special  
CC methods. The present invention is a polypeptide encoded by a full length  
CC human cDNA of the invention. Note: The sequence data for this patent did  
CC not form part of the printed specification, but was obtained in CD-ROM

CC format directly from EPO  
XX  
SQ Sequence 263 AA;

Query Match 99.0%; Score 1374; DB 4; Length 263;  
Best Local Similarity 99.2%; Pred. No. 1.3e-117;  
Matches 260; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 LLAWQAFVLSNMLLAAYGSGGCFWNGHLYREDQTSAPGLRCLNWLDAQSLASAPV 60  
DB 2 LLAWQAFVLSNMLLAAYGSGGCFWNGHLYREDQTSAPGLRCLNWLDAQSLASAPV 61

QY 61 SGAGNHSYCRNPDEDPGPGWCYVSGEAGVPEKPCEDLRCPETTSQALPAFTTEIQEASE 120  
DB 62 SGAGNHSYCRNPDEDPGPGWCYVSGEAGVPEKPCEDLRCPETTSQALPAFTTEIQEASE 121

QY 121 GPGADEVQVFPANALPARSEAAAVQPVIGISQVRMNSKEKKDLGTLGYVLGTTMMVII 180  
DB 122 GPGADEVQVFPANALPARSEAAAVQPVIGISQVRMNSKEKKDLGTLGYVLGTTMMVII 181

QY 181 IATGAGIILGYSGKGLKEQHDQKVCEREMQRTITPLSAFTNPTCEIVDEKTVVHTS 240  
DB 182 IATGAGIILGYSGKGLKEQHDQKVCEREMQRTITPLSAFTNPTCEIVDEKTVVHTS 241

QY 241 QTFVDPQEGSTPLMGQAGTPGA 262  
DB 242 QTFVDPQEGSTPLMGQAGTPGA 263

RESULT 8  
AAY05220  
ID AAY05220 standard; protein; 286 AA.  
XX  
AC AAY05220;  
XX  
DT 17-JUN-1999 (first entry)  
XX  
DE Kringle1 protein sequence.  
XX  
XX Kringle1; cancer; inflammation; autoimmunity; rheumatoid arthritis; AIDS;  
KW CNS inflammation; cerebellar degeneration; Alzheimer's disease; asthma;  
KW Parkinson's disease; multiple sclerosis; amyotrophic lateral sclerosis;  
KW neurological abnormality; ischaemia reperfusion injury; ischaemic injury;  
KW cardiovascular disease; kidney disease; liver disease; aplastic anaemia;  
KW myocardial infarction; hypotension; hypertension; allergy; infection;  
KW myelodysplastic syndrome; haematologic abnormality; diagnosis; therapy;  
KW male pattern baldness.  
XX  
OS Homo sapiens.  
XX  
PN WO9911788-A1.  
XX  
PD 11-MAR-1999.  
XX  
XX  
PF 02-SEP-1998; 98WO-US018270.  
XX  
PR 02-SEP-1997; 97US-0056032P.  
PR 01-SEP-1998; 98US-00144889.  
XX  
XX (SMK ) SMITHKLINE BEECHAM CORP.  
XX  
XX Albone EF, Kikly KK;  
PI  
XX  
XX WPI; 1999-214707/18.  
DR N-PSDB; AAX28355.  
XX  
XX New kringle1 polypeptides and polynucleotides.  
PT  
XX  
XX Claim 14; Page 33; 42pp; English.  
PS  
XX This sequence is a Kringle1 polypeptide of the invention. The kringle1  
CC polypeptides (I) are used to screen for agonists and antagonists.  
CC Agonists are used to treat subjects in need of enhanced activity or

expression of (I). Antagonists are used to treat subjects having need to inhibit the activity or expression of (I). The methods can be used to treat conditions such as cancer, inflammation, autoimmunity, allergy, asthma, rheumatoid arthritis, CNS inflammation, cerebellar degeneration, Alzheimer's disease, Parkinson's disease, multiple sclerosis, amyotrophic lateral sclerosis, head injury damage and other neurological abnormalities, ischaemia reperfusion injury, cardiovascular disease, kidney disease, liver disease, ischaemic injury, myocardial infarction, hypotension, hypertension, AIDS, myelodysplastic syndromes and other haematologic abnormalities, aplastic anaemia, male pattern baldness, and bacterial, fungal, protozoan and viral infections. The kringle1 polypeptides may also be used to generate antibodies. Determining the presence or absence of mutations in, and analysing for the presence or absence of expression of, kringle1 polynucleotides can be used to diagnose a disease or susceptibility to a disease related to expression or activity of kringle1 proteins. The polynucleotides may also be used for chromosome identification, and mapping

XX Sequence 286 AA;

Query Match 83.8%; Score 1162.5; DB 2; Length 286;  
Best Local Similarity 87.3%; Pred. No. 3.8e-98;  
Matches 226; Conservative 5; Mismatches 27; Indels 1; Gaps 1;

QY 1 LLAWQAFVLSNMLLAAYSGGCFWNGHLYREDQTSAPGLRCLNWLDAQSLASAPV 60  
DB 2 LLAWQAFVLSNMLLAAYSGGCFWNGHLYREDQTSAPGLRCLNWLDAQSLASAPV 61  
QY 61 SGAGNHSYCRNPDPRGRCYVSGAGVPEKPCEDLRCPETTSQALPAF-TTIOEAS 119  
DB 62 SGADNHSYCRNPDPRGRCYVSGAGVPEKPCEDLRCPETTSQALPAFPHDRNSRKL 121  
QY 120 EGCADDEVQVFAPANALPARSEAAVQPVIGISQVRNMSKEKDLGLGVLTMMVI 179  
DB 122 EGCADDEVQVFAPANALPARSEAAVQPVIGISQVRNMSKEKDLGLGVLTMMVI 181  
QY 180 IIAIGAGIILGYSGKDLKEQDKVCEREMQRTITPLSFAFTNPTCEIVDEKTVVHT 239  
DB 182 IIAIGAGIILGYSGKDLKEQDKVCEREMQRTITPLSFAFTNPTCEIVDEKTVVHT 241  
QY 240 SQTVPDQEGSTPLMGAG 258  
DB 242 SQTVPDQEGSTPLMGAG 260

RESULT 9  
ABR40487  
ID ABR40487 standard; protein; 146 AA.

XX ABR40487;

XX 13-JUN-2003 (first entry)

XX Human secreted protein #SEQ ID 237.

XX Human; secreted protein; anti-HIV; nootropic; neuroprotective;  
XX antiangiogenic; immunosuppressive; immunomodulator; cytostatic; cardiant;  
XX hepatotropic; antiinflammatory; antiallergic; antidiabetic;  
XX gastrointestinal; antifertility; nephrotropic; virucide; hypotensive;  
XX vasotropic; dermatological; osteopathic; antiarthritic; antiparkinsonian;  
XX antiaesthetic; antipsoriatic; cerebroprotective; antibacterial;  
XX fungicide; antiparasitic; protein therapy; gene therapy; cancer;  
XX hyperproliferative disorder; leukaemia; autoimmune disorder;  
XX immunodeficiency; HIV; infertility; uterine disorders; endometriosis;  
XX amenorrhea; ocular disorder; neurological disorder; wound healing;  
XX Huntington's disease; gastrointestinal disorder; inflammatory disease;  
XX Crohn's disease; kidney disorder; renal failure; diabetes; obesity;  
XX cirrhosis; hepatitis; cardiovascular disorder; angina; osteoarthritis;  
XX rheumatoid arthritis; psoriasis; multiple sclerosis; asthma;  
XX respiratory disorder; infectious disease; chromosome identification;  
XX food additive; nutrition.

OS Homo sapiens.

XX PN WO200268628-A1.  
XX PD 06-SEP-2002.  
XX 21-FEB-2002; 2002WO-US005301.  
XX 23-FEB-2001; 2001US-0270625P.  
PR 12-JUL-2001; 2001US-0304417P.  
XX (HUMA-) HUMAN GENOME SCI INC.  
XX Rosen CA, Komatsulis G, Baker KP, Fiscella M, Moore PA, Wei P;  
PI Duan DR, Shi Y, Gupta R;  
XX WPI; 2002-750417/81.  
DR N-PSDB; AB282542.  
XX  
XX New human secreted proteins and nucleic acids, useful for preventing,  
XX treating, or ameliorating a medical condition, e.g. cancer, autoimmune  
XX disorder, HIV, Parkinson's disease, Alzheimer's disease, diabetes,  
XX obesity or cirrhosis.  
XX  
XX Claim 11; Page 800; 873pp; English.

XX The invention relates to novel human secreted proteins and the genes  
XX encoding them. Genes and proteins of the invention may be useful for  
XX preventing, treating or ameliorating medical conditions e.g. by protein  
XX or gene therapy. These conditions include cancer and hyperproliferative  
XX disorders, immune cell proliferative disorders (e.g. leukaemia),  
XX autoimmune disorders, immunodeficiencies (e.g. HIV infection),  
XX infertility, placental and uterine disorders (e.g. endometriosis),  
XX amenorrhea, ocular disorders, neurological disorders (e.g. Huntington's  
XX disease), wound healing, gastrointestinal system disorders, particularly  
XX inflammatory diseases (e.g. Crohn's disease), kidney disorders (e.g.  
XX renal failure), diabetes, obesity, cirrhosis, hepatitis, cardiovascular  
XX disorders (e.g. angina), rheumatoid arthritis, osteoarthritis, psoriasis,  
XX multiple sclerosis, respiratory disorders (e.g. asthma), or infectious  
XX diseases caused by bacterial, parasitic viral or fungal agents. The  
XX nucleic acids are also useful for chromosome identification, radiation  
XX hybrid mapping or long-range restriction mapping. The polypeptide,  
XX polynucleotide, agonist or antagonist may also be used as a food additive  
XX or preservative to increase or decrease storage capabilities, fat content  
XX or other nutritional components. The sequences given in records ABR40409-  
XX ABR40590 and AB282464-AB282611 represent human secreted proteins and the  
XX genes encoding them

XX Sequence 146 AA;

Query Match 53.8%; Score 747; DB 5; Length 146;  
Best Local Similarity 94.5%; Pred. No. 2e-60;  
Matches 137; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

QY 1 LLAWQAFVLSNMLLAAYSGGCFWNGHLYREDQTSAPGLRCLNWLDAQSLASAPV 60  
DB 2 LLAWQAFVLSNMLLAAYSGGCFWNGHLYREDQTSAPGLRCLNWLDAQSLASAPV 61  
QY 61 SGAGNHSYCRNPDPRGRCYVSGAGVPEKPCEDLRCPETTSQALPAFTTIOEASE 120  
DB 62 SGAGNHSYCRNPDPRGRCYVSGAGVPEKPCEDLRCPETTSQALPAFTTIOEASE 121  
QY 121 GPCADDEVQVFAPANALPARSEAAV 145  
DB 122 GPCADDEVQVFAPATPYPLGSAVAV 146

RESULT 10

ABR40561  
ID ABR40561 standard; protein; 146 AA.

XX ABR40561;

XX 13-JUN-2003 (first entry)

XX DE Human secreted protein #SEQ ID 311.  
 XX KW Human; secreted protein; anti-HIV; neurotropic; neuroprotective;  
 KW antiangiogenic; immunosuppressive; immunomodulator; cytostatic; cardiac;  
 KW hepatotropic; antiinflammatory; antiallergic; antidiabetic;  
 KW gastrointestinal; antiinfertility; nephrotropic; virucide; hypotensive;  
 KW vasotropic; dermatological; osteopathic; antiarthritic; antiparkinsonian;  
 KW angiodysplastic; antiparasitic; cerebroprotective; antibacterial;  
 KW fungicidal; antiparasitic; protein therapy; gene therapy; cancer;  
 KW hyperproliferative disorder; leukaemia; autoimmune disorder;  
 KW immunodeficiency; HIV; infertility; uterine disorders; endometriosis;  
 KW amenorrhea; ocular disorder; neurological disorder; wound healing;  
 KW Huntington's disease; gastrointestinal disorder; inflammatory disease;  
 KW Crohn's disease; kidney disorder; renal failure; diabetes; obesity;  
 KW cirrhosis; hepatitis; cardiovascular disorder; angina; osteoarthritis;  
 KW rheumatoid arthritis; psoriasis; multiple sclerosis; asthma;  
 KW respiratory disorder; infectious disease; chromosome identification;  
 KW food additive; nutrition.  
 XX OS Homo sapiens.  
 XX PN WO200268628-A1.  
 XX PD 06-SEP-2002.  
 XX PF 21-FEB-2002; 2002WO-US005301.  
 XX PR 23-FEB-2001; 2001US-0270625P.  
 XX PR 12-JUL-2001; 2001US-0304417P.  
 XX PA (HUMA-) HUMAN GENOME SCI INC.  
 XX PI Rosen CA, Komatsoulis G, Baker KP, Piscella M, Moore PA, Wei P;  
 PI Duan DR, Shi Y, Gupta R;  
 XX DR WPI; 2002-750417/81.  
 XX PT New human secreted proteins and nucleic acids, useful for preventing,  
 PT treating, or ameliorating a medical condition, e.g. cancer, autoimmune  
 PT disorder, HIV, Parkinson's disease, Alzheimer's disease, diabetes,  
 PT obesity or cirrhosis.  
 XX PS Claim 11; Page 850; 873pp; English.  
 XX CC The invention relates to novel human secreted proteins and the genes  
 CC encoding them. Genes and proteins of the invention may be useful for  
 CC preventing, treating or ameliorating medical conditions e.g. by protein  
 CC or gene therapy. These conditions include cancer and hyperproliferative  
 CC disorders, immune cell proliferative disorders (e.g. leukaemia),  
 CC autoimmune disorders, immunodeficiencies (e.g. HIV infection),  
 CC infertility, placental and uterine disorders (e.g. endometriosis),  
 CC amenorrhea, ocular disorders, neurological disorders (e.g. Huntington's  
 CC disease), wound healing, gastrointestinal system disorders, particularly  
 CC inflammatory diseases (e.g. Crohn's disease), kidney disorders (e.g.  
 CC renal failure), diabetes, obesity, cirrhosis, hepatitis, cardiovascular  
 CC disorders (e.g. angina), rheumatoid arthritis, osteoarthritis, psoriasis,  
 CC multiple sclerosis, respiratory disorders (e.g. asthma), or infectious  
 CC diseases caused by bacterial, parasitic, viral or fungal agents. The  
 CC nucleic acids are also useful for chromosome identification, radiation  
 CC hybrid mapping or long-range restriction mapping. The polypeptide,  
 CC polynucleotide, agonist or antagonist may also be used as a food additive  
 CC or preservative to increase or decrease storage capabilities, fat content  
 CC or other nutritional components. The sequences given in records ABR40409-  
 CC ABR40590 and ABR282464-ABR282611 represent human secreted proteins and the  
 CC genes encoding them  
 XX SQ Sequence 146 AA;

Query Match 53.8%; Score 747; DB 5; Length 146;  
 Best Local Similarity 94.5%; Pred. No. 2e-60;  
 Matches 137; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

Oy 1 LLAWOAFVSNMLAEAYGSGCFWDMHLYRSDQTSAPGLRCLNWLDAQSGIASAPV 60  
 Db 2 LLAWOAFVSNMLAEAYGSGCFWDMHLYRSDQTSAPGLRCLNWLDAQSGIASAPV 61  
 Oy 61 SGAGNSHCNRPDPRGVCWCVSGEAGVPEKPCEDLRCPCETTSOALPATTETIQEASE 120  
 Db 62 SGAGNSHCNRPDPRGVCWCVSGEAGVPEKPCEDLRCPCETTSOALPATTETIQEASE 121  
 Oy 121 GPGADEVQVFPANALPARSEAAV 145  
 Db 122 GPGADEVQVFPATXPPLGSGXAV 146  
 RESULT 11  
 ABR42624  
 ID ABR42624 standard; protein; 81 AA.  
 XX ABR42624;  
 AC ABR42624;  
 XX 26-AUG-2003 (first entry)  
 DT XX  
 XX Human kringle containing protein.  
 DE XX  
 XX Human; abrogen; kringle; angiogenesis; inhibitor; tumour; metastasis;  
 KW cytosolic; gene therapy; expressed sequence tag; EST.  
 XX OS Homo sapiens.  
 XX Key Location/Qualifiers  
 FT Misc-difference 4 /note= "the identity of this residue is unclear in Fig 2"  
 FT Misc-difference 6 /note= "the identity of this residue is unclear in Fig 2"  
 FT Misc-difference 7 /note= "the identity of this residue is unclear in Fig 2"  
 FT Misc-difference 16 /note= "the identity of this residue is unclear in Fig 2"  
 FT Misc-difference 29 /note= "the identity of this residue is unclear in Fig 2"  
 FT Misc-difference 32 /note= "the identity of this residue is unclear in Fig 2"  
 FT Misc-difference 46 /note= "the identity of this residue is unclear in Fig 2"  
 FT Misc-difference 50..56 /note= "the identity of this residue is unclear in Fig 2"  
 FT Domain /note= "kringle domain"  
 FT Misc-difference 57 /note= "the identity of this residue is unclear in Fig 2"  
 FT Misc-difference 62 /note= "the identity of this residue is unclear in Fig 2"  
 FT Misc-difference 70 /note= "the identity of this residue is unclear in Fig 2"  
 FT Misc-difference 72 /note= "the identity of this residue is unclear in Fig 2"  
 FT Misc-difference 77 /note= "the identity of this residue is unclear in Fig 2"  
 XX WO2003042354-A2.  
 XX 22-MAY-2003.  
 XX 04-SEP-2002; 2002WO-US027885.  
 XX 04-SEP-2001; 2001US-0316300P.  
 XX (AVET ) AVENTIS PHARM INC.  
 XX Nesbit M, Fong TC, Brockstedt D;  
 XX WPI; 2003-449566/42.  
 XX New abrogen polypeptide, useful for treating an angiogenesis related  
 PT diseases e.g. tumor metastasis.  
 PT

PS Disclosure; Fig 2; 95pp; English.

XX The present sequence is the protein sequence of a hypothetical kringle-  
CC containing protein encoded by an expressed sequence tag. The invention  
CC relates to novel abrogen polypeptides that are derived from kringle-  
CC containing proteins. The abrogens are potent inhibitors of endothelial  
CC proliferation and angiogenesis. They are capable of inhibiting or  
CC reducing cell proliferation induced by both basic fibroblast growth  
CC factor and vascular endothelial growth factor in a specific endothelial  
CC cell proliferation assay. Vectors that expressed abrogen polypeptides in  
CC vivo were shown to reduce tumour metastasis in 2 lung cancer models. The  
CC invention provides abrogen polypeptides and polynucleotides, and methods  
CC of using these to treat an angiogenesis-related disease or disorder, e.g.  
CC tumour metastasis (claimed)

XX Sequence 81 AA;  
SQ

Query Match 27.7%; Score 385; DB 7; Length 81;  
Best Local Similarity 82.7%; Pred. No. 1.5e-27;  
Matches 67; Conservative 0; Mismatches 14; Indels 0; Gaps 0;

QY 22 GGCFTDNGHLYREDQTSFAPGLRCLNWLDAQSLASAPVSGAGNHSYCRNPDPRGPC 81  
DB 1 GGCXWXXGHLREDQXSFAGLRLNWLXQAQXGLASAPVSGAGNHSYCRNPDPRGPC 60

QY 82 YVSGEAGVPEKPCEDLRCE 102  
DB 61 YXSGEAGVPEKPCEDXRCE 81

RESULT 12  
AAM1800  
ID AAM1800 standard; protein; 66 AA.  
AC AAM1800;  
XX  
XX 12-OCT-2001 (first entry)  
XX  
XX Peptide #5234 encoded by probe for measuring cervical gene expression.  
XX Probe; human; microarray; gene expression; cervical epithelial cell;  
XX cervical cancer.  
XX Homo sapiens.  
XX  
XX WO200157278-A2.  
XX  
XX 09-AUG-2001.  
XX  
XX 30-JAN-2001; 2001WO-US0000670.  
XX  
XX 04-FEB-2000; 2000US-0180312P.  
XX 26-MAY-2000; 2000US-0207456P.  
XX 30-JUN-2000; 2000US-00608408.  
XX 03-AUG-2000; 2000US-00632366.  
XX 21-SEP-2000; 2000US-0234687P.  
XX 27-SEP-2000; 2000US-0236359P.  
XX 04-OCT-2000; 2000GB-00024263.  
XX (MOLE-) MOLECULAR DYNAMICS INC.  
XX  
XX Penn SG, Hanzel DK, Chen W, Rank DR;  
XX  
XX WPI; 2001-489901/53.  
XX  
XX Human genome-derived single exon nucleic acid probes useful for analyzing  
XX gene expression in human cervical epithelial cells.  
XX  
XX Claim 27; SEQ ID NO 23626; 487pp; English.  
XX  
XX The present invention relates to human single exon nucleic acid probes  
XX (SENPs: see AAM10068-AA128459). The present sequence is a peptide encoded  
XX by one such probe. The SENPs are derived from human HeLa cells. The SENPs

CC can be used to produce a single exon microarray, which can be used for  
CC measuring human gene expression in a sample derived from human cervical  
CC epithelial cells. By measuring gene expression, the probes are therefore  
CC useful in grading and/or staging of diseases of the cervix, notably  
CC cervical cancer. Note: The sequence data for this patent did not form  
CC part of the printed specification, but was obtained in electronic format  
CC directly from WIPO at ftp.wipo.int/pub/published\_pct\_sequences

XX Sequence 66 AA;  
SQ

Query Match 23.2%; Score 322; DB 4; Length 66;  
Best Local Similarity 100.0%; Pred. No. 7.1e-22;  
Matches 66; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 103 TTSQLPAPFTTEIQEASGEGDEVQVFAFANALPARSEAAAVQPVIGISQVRMNSKEK 162  
DB 1 TTSQLPAPFTTEIQEASGEGDEVQVFAFANALPARSEAAAVQPVIGISQVRMNSKEK 60

QY 163 KDLGTL 168  
DB 61 KDLGTL 66

RESULT 13  
ABB37905  
ID ABB37905 standard; peptide; 66 AA.  
XX  
XX ABB37905;  
XX  
XX 04-FEB-2002 (first entry)  
XX  
XX Peptide #5411 encoded by human foetal liver single exon probe.  
XX Human; foetal liver; gene expression; single exon nucleic acid probe.  
XX Homo sapiens.  
XX  
XX WO200157277-A2.  
XX  
XX 09-AUG-2001.  
XX  
XX 30-JAN-2001; 2001WO-US0000669.  
XX  
XX 04-FEB-2000; 2000US-0180312P.  
XX 26-MAY-2000; 2000US-0207456P.  
XX 30-JUN-2000; 2000US-00608408.  
XX 03-AUG-2000; 2000US-00632366.  
XX 21-SEP-2000; 2000US-0234687P.  
XX 27-SEP-2000; 2000US-0236359P.  
XX 04-OCT-2000; 2000GB-00024263.  
XX (MOLE-) MOLECULAR DYNAMICS INC.  
XX  
XX Penn SG, Hanzel DK, Chen W, Rank DR;  
XX  
XX WPI; 2001-483447/52.  
XX  
XX Human genome-derived single exon nucleic acid probes useful for analyzing  
XX gene expression in human fetal liver.  
XX  
XX Claim 27; SEQ ID NO 30540; 639pp + Sequence Listing; English.  
XX  
XX The invention relates to a single exon nucleic acid probe for measuring  
XX human gene expression in a sample derived from human foetal liver. The  
XX single exon nucleic acid probes may be used for predicting, measuring and  
XX displaying gene expression in samples derived from human fetal liver. The  
XX present sequence is a peptide encoded by a single exon nucleic acid probe  
XX of the invention. Note: The sequence data for this patent did not form  
XX part of the printed specification, but was obtained in electronic format  
XX directly from WIPO at ftp.wipo.int/pub/published\_pct\_sequences

XX Sequence 66 AA;  
SQ

Query Match 23.2%; Score 322; DB 4; Length 66;  
Best Local Similarity 100.0%; Pred. No. 7.1e-22;  
Matches 66; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 103 TTSCALPAFTTEIQEASEGPGADEVQVFAPANALPARSEAAAVQVIGISQVRMNSKEK 162  
DB 1 TTSCALPAFTTEIQEASEGPGADEVQVFAPANALPARSEAAAVQVIGISQVRMNSKEK 60  
QY 163 KDLGTL 168  
DB 61 KDLGTL 66

RESULT 14  
AAM31314  
ID AAM31314 standard; protein; 66 AA.  
XX AAM31314;  
AC AAM31314;  
DT 17-OCT-2001 (first entry)  
DE Peptide #531 encoded by probe for measuring placental gene expression.  
XX Probe; microarray; human; placenta; antenatal diagnosis;  
KW genetic disorder.  
OS Homo sapiens.  
XX WO200157272-A2.  
PN 09-AUG-2001.  
PD 30-JAN-2001; 2001WO-US000663.  
PF 04-FEB-2000; 2000US-0180312P.  
PR 26-MAY-2000; 2000US-0207456P.  
PR 30-JUN-2000; 2000US-00608408.  
PR 03-AUG-2000; 2000US-00632366.  
PR 21-SEP-2000; 2000US-0234687P.  
PR 27-SEP-2000; 2000US-0236359P.  
PR 04-OCT-2000; 2000GB-00024263.  
XX (MOLE-) MOLECULAR DYNAMICS INC.  
XX Penn SG, Hanzel DK, Chen W, Rank DR;  
XX WPI; 2001-488897/53.  
XX Human genome-derived single exon nucleic acid probes useful for analyzing  
PT gene expression in human placenta.  
XX Claim 27; SEQ ID NO 31583; 654pp; English.

The present invention relates to single exon nucleic acid probes (SENP:  
see A131315-AA157546). The present sequence is a peptide encoded by one  
such probe. The probes are useful for producing a microarray for  
predicting, measuring and displaying gene expression in samples derived  
from human placenta. The probes are useful for antenatal diagnosis of  
human genetic disorders  
XX Sequence 66 AA;

Query Match 23.2%; Score 322; DB 4; Length 66;  
Best Local Similarity 100.0%; Pred. No. 7.1e-22;  
Matches 66; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 103 TTSCALPAFTTEIQEASEGPGADEVQVFAPANALPARSEAAAVQVIGISQVRMNSKEK 162  
DB 1 TTSCALPAFTTEIQEASEGPGADEVQVFAPANALPARSEAAAVQVIGISQVRMNSKEK 60  
QY 163 KDLGTL 168  
DB 61 KDLGTL 66

RESULT 15  
ABB23159  
ID ABB23159 standard; protein; 66 AA.  
XX ABB23159;  
AC ABB23159;  
DT 23-JAN-2002 (first entry)  
DE Protein #518 encoded by probe for measuring heart cell gene expression.  
XX Human; gene expression; heart; microarray; vascular system;  
KW cardiovascular disease; hypertension; cardiac arrhythmia;  
KW congenital heart disease.  
XX Homo sapiens.  
XX WO200157274-A2.  
PN 09-AUG-2001.  
PD 30-JAN-2001; 2001WO-US000666.  
PF 04-FEB-2000; 2000US-0180312P.  
PR 26-MAY-2000; 2000US-0207456P.  
PR 30-JUN-2000; 2000US-00608408.  
PR 03-AUG-2000; 2000US-00632366.  
PR 21-SEP-2000; 2000US-0234687P.  
PR 27-SEP-2000; 2000US-0236359P.  
PR 04-OCT-2000; 2000GB-00024263.  
XX (MOLE-) MOLECULAR DYNAMICS INC.  
XX Penn SG, Hanzel DK, Chen W, Rank DR;  
XX WPI; 2001-488899/53.  
XX Single exon nucleic acid probes for analyzing gene expression in human  
PT hearts.  
XX Claim 15; SEQ ID NO 24929; 530pp; English.

The present invention relates to single exon nucleic acid probes for  
measuring human gene expression in a sample derived from human heart (see  
ABA21535-ABA41305). The present sequence is a protein encoded by one such  
probe. The probes may be used for predicting, measuring and displaying  
gene expression in samples derived from the human heart via microarrays.  
By measuring gene expression, the probes are useful for predicting,  
diagnosing, grading, staging, monitoring and prognosing diseases of the  
human heart and vascular system e.g. cardiovascular disease,  
hypertension, cardiac arrhythmias and congenital heart disease. Note: The  
sequence data for this patent did not form part of the printed  
specification, but was obtained in electronic format directly from WIPO  
at ftp.wipo.int/pub/published\_pct\_sequences  
XX Sequence 66 AA;

Query Match 23.2%; Score 322; DB 4; Length 66;  
Best Local Similarity 100.0%; Pred. No. 7.1e-22;  
Matches 66; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 103 TTSCALPAFTTEIQEASEGPGADEVQVFAPANALPARSEAAAVQVIGISQVRMNSKEK 162  
DB 1 TTSCALPAFTTEIQEASEGPGADEVQVFAPANALPARSEAAAVQVIGISQVRMNSKEK 60  
QY 163 KDLGTL 168  
DB 61 KDLGTL 66

Search completed: March 17, 2004, 07:03:42  
Job time : 68.9986 secs

---

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 18, 2004, 13:12:33 ; Search time 47.6043 Seconds  
(without alignments)  
1425.213 Million cell updates/sec

Title: US-10-057-951-2\_COPY\_2\_263  
Perfect score: 1388  
Sequence: 1 LLAWQAFVLSNMLLAAYG.....PVDPQEGSTPLMGQGTGGA 262

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1049977 seqs, 258955339 residues

Total number of hits satisfying chosen parameters: 1049977

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Published Applications AA:\*\*

1: /cgn2\_6/ptodata/2/pubpaa/US07\_PUBCOMB.pep:\*  
2: /cgn2\_6/ptodata/2/pubpaa/PCT\_NEW\_PUB.pep:\*  
3: /cgn2\_6/ptodata/2/pubpaa/US03\_NEW\_PUB.pep:\*  
4: /cgn2\_6/ptodata/2/pubpaa/US06\_PUBCOMB.pep:\*  
5: /cgn2\_6/ptodata/2/pubpaa/US07\_NEW\_PUB.pep:\*  
6: /cgn2\_6/ptodata/2/pubpaa/PTUS\_PUBCOMB.pep:\*  
7: /cgn2\_6/ptodata/2/pubpaa/US08\_NEW\_PUB.pep:\*  
8: /cgn2\_6/ptodata/2/pubpaa/US08\_PUBCOMB.pep:\*  
9: /cgn2\_6/ptodata/2/pubpaa/US09A\_PUBCOMB.pep:\*  
10: /cgn2\_6/ptodata/2/pubpaa/US09B\_PUBCOMB.pep:\*  
11: /cgn2\_6/ptodata/2/pubpaa/US09C\_PUBCOMB.pep:\*  
12: /cgn2\_6/ptodata/2/pubpaa/US09\_NEW\_PUB.pep:\*  
13: /cgn2\_6/ptodata/2/pubpaa/US10A\_PUBCOMB.pep:\*  
14: /cgn2\_6/ptodata/2/pubpaa/US10B\_PUBCOMB.pep:\*  
15: /cgn2\_6/ptodata/2/pubpaa/US10C\_PUBCOMB.pep:\*  
16: /cgn2\_6/ptodata/2/pubpaa/US10\_NEW\_PUB.pep:\*  
17: /cgn2\_6/ptodata/2/pubpaa/US60\_NEW\_PUB.pep:\*  
18: /cgn2\_6/ptodata/2/pubpaa/US60\_PUBCOMB.pep:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query Match	Score	Length	ID	Description
1	1388	100.0	263	9	US-09-084-491A-2
2	1388	100.0	263	13	US-10-102-704-2
3	1388	100.0	263	13	US-10-057-951-2
4	1385	99.8	263	14	US-10-210-951-44
5	1385	99.8	263	14	US-10-211-884-44
6	322	23.2	66	9	US-09-864-761-38457
7	160.5	11.6	650	15	US-10-401-077-1
8	154.5	11.1	527	10	US-09-987-457-18
9	154.5	11.1	527	10	US-09-987-455-19
10	154.5	11.1	527	15	US-10-360-101-203
11	154.5	11.1	562	9	US-09-969-271-7
12	154.5	11.1	562	9	US-09-974-298-145
13	154.5	11.1	562	12	US-10-411-037-26
14	154.5	11.1	562	14	US-10-193-656-8
15	154.5	11.1	562	14	US-10-443-701-4
					Sequence 2, Appli
					Sequence 2, Appli
					Sequence 2, Appli
					Sequence 44, Appli
					Sequence 44, Appli
					Sequence 38457, A
					Sequence 1, Appli
					Sequence 18, Appli
					Sequence 19, Appli
					Sequence 203, App
					Sequence 7, Appli
					Sequence 145, App
					Sequence 26, Appli
					Sequence 8, Appli
					Sequence 4, Appli

US-09-084-491A-2  
Sequence 2, Application US/09084491A  
Patent No. US20020061576A1  
GENERAL INFORMATION:  
APPLICANT: WOORE, PAUL A.  
APPLICANT: RUBEN, STEVEN M.  
APPLICANT: EBER, REINHARD  
TITLE OF INVENTION: TISSUE PLASMINOGEN ACTIVATOR-LIKE PROTEASE  
NUMBER OF SEQUENCES: 16  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: HUMAN GENOME SCIENCES, INC.  
STREET: 9410 KEY WEST AVENUE  
CITY: ROCKVILLE  
STATE: MD  
COUNTRY: USA  
ZIP: 20850  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/084,491A  
FILING DATE: 27-MAY-1998  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: BROOKES, ANDERS A.  
REGISTRATION NUMBER: 36,373  
REFERENCE/DOCKET NUMBER: PF378  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (301) 309-8504  
TELEFAX: (301) 309-8439  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 263 amino acids  
TYPE: amino acid  
TOPOLOGY: linear

ALIGNMENTS

RESULT 1

US-09-084-491A-2

Sequence 2, Application US/09084491A

Patent No. US20020061576A1

GENERAL INFORMATION:

APPLICANT: WOORE, PAUL A.

APPLICANT: RUBEN, STEVEN M.

APPLICANT: EBER, REINHARD

TITLE OF INVENTION: TISSUE PLASMINOGEN ACTIVATOR-LIKE PROTEASE

NUMBER OF SEQUENCES: 16

CORRESPONDENCE ADDRESS:

ADDRESSEE: HUMAN GENOME SCIENCES, INC.

STREET: 9410 KEY WEST AVENUE

CITY: ROCKVILLE

STATE: MD

COUNTRY: USA

ZIP: 20850

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/084,491A

FILING DATE: 27-MAY-1998

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: BROOKES, ANDERS A.

REGISTRATION NUMBER: 36,373

REFERENCE/DOCKET NUMBER: PF378

TELECOMMUNICATION INFORMATION:

TELEPHONE: (301) 309-8504

TELEFAX: (301) 309-8439

INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:

LENGTH: 263 amino acids

TYPE: amino acid

TOPOLOGY: linear

```
MOLECULE TYPE: protein
US-09-064-491A-2

Query Match      100.0%; Score 1388; DB 9; Length 263;
Best Local Similarity 100.0%; Pred. No. 3.4e-125;
Matches 262; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LLAWQAFVLSNMLLAAYGSGGCFWNGHLYREDQTSAPGLRCLNWLDAQSLASAPV 60
DB 2 LLAWQAFVLSNMLLAAYGSGGCFWNGHLYREDQTSAPGLRCLNWLDAQSLASAPV 61

QY 61 SGAGNHSYCRNPDEDPGPMWCVYSGEAGVPEKPCEDLRCPETTSQALPAFTTEIOEASE 120
DB 62 SGAGNHSYCRNPDEDPGPMWCVYSGEAGVPEKPCEDLRCPETTSQALPAFTTEIOEASE 121

QY 121 GPGADEVQVFAPANALPARSEAAAVQPVIGISQVRVNSKEKKDLGTLGYVLGITMNVII 180
DB 122 GPGADEVQVFAPANALPARSEAAAVQPVIGISQVRVNSKEKKDLGTLGYVLGITMNVII 181

QY 181 IAIGAGIILGYSYKRGKDLKEQHDQKVCEREMQRIITPLSAFTNPTCEIVDEKTVVWHTS 240
DB 182 IAIGAGIILGYSYKRGKDLKEQHDQKVCEREMQRIITPLSAFTNPTCEIVDEKTVVWHTS 241

QY 241 QTPVDPQEGSTPLMGQAGTPGA 262
DB 242 QTPVDPQEGSTPLMGQAGTPGA 263

RESULT 2
US-10-102-704-2
; Sequence 2, Application US/10102704
; Publication No. US20020164768A1
; GENERAL INFORMATION:
; APPLICANT: Moore et al.
; TITLE OF INVENTION: Tissue Plasminogen Activator-Like Protein
; FILE REFERENCE: PF378C1
; CURRENT APPLICATION NUMBER: US/10/102,704
; CURRENT FILING DATE: 2002-03-22
; PRIOR APPLICATION NUMBER: 09/084,491
; PRIOR FILING DATE: 1998-05-27
; PRIOR APPLICATION NUMBER: 60/048,000
; PRIOR FILING DATE: 1997-05-28
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: Patent in version 3.1
; SEQ ID NO 2
; LENGTH: 263
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-102-704-2

Query Match      100.0%; Score 1388; DB 13; Length 263;
Best Local Similarity 100.0%; Pred. No. 3.4e-125;
Matches 262; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LLAWQAFVLSNMLLAAYGSGGCFWNGHLYREDQTSAPGLRCLNWLDAQSLASAPV 60
DB 2 LLAWQAFVLSNMLLAAYGSGGCFWNGHLYREDQTSAPGLRCLNWLDAQSLASAPV 61

QY 61 SGAGNHSYCRNPDEDPGPMWCVYSGEAGVPEKPCEDLRCPETTSQALPAFTTEIOEASE 120
DB 62 SGAGNHSYCRNPDEDPGPMWCVYSGEAGVPEKPCEDLRCPETTSQALPAFTTEIOEASE 121

QY 121 GPGADEVQVFAPANALPARSEAAAVQPVIGISQVRVNSKEKKDLGTLGYVLGITMNVII 180
DB 122 GPGADEVQVFAPANALPARSEAAAVQPVIGISQVRVNSKEKKDLGTLGYVLGITMNVII 181

QY 181 IAIGAGIILGYSYKRGKDLKEQHDQKVCEREMQRIITPLSAFTNPTCEIVDEKTVVWHTS 240
DB 182 IAIGAGIILGYSYKRGKDLKEQHDQKVCEREMQRIITPLSAFTNPTCEIVDEKTVVWHTS 241

QY 241 QTPVDPQEGSTPLMGQAGTPGA 262
DB 242 QTPVDPQEGSTPLMGQAGTPGA 263

RESULT 3
US-10-057-951-2
; Sequence 2, Application US/10057951
; Publication No. US20020177213A1
; GENERAL INFORMATION:
; APPLICANT: Moore et al.
; TITLE OF INVENTION: Tissue Plasminogen Activator-Like Protease
; FILE REFERENCE: PF378P1
; CURRENT APPLICATION NUMBER: US/10/057,951
; CURRENT FILING DATE: 2002-01-23
; PRIOR APPLICATION NUMBER: US 09/411,977
; PRIOR FILING DATE: 1999-10-04
; PRIOR APPLICATION NUMBER: US 09/084,491
; PRIOR FILING DATE: 1998-05-27
; PRIOR APPLICATION NUMBER: US 60/048,000
; PRIOR FILING DATE: 1997-05-28
; NUMBER OF SEQ ID NOS: 31
; SOFTWARE: Patent in ver. 3.1
; SEQ ID NO 2
; LENGTH: 263
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-057-951-2

Query Match      100.0%; Score 1388; DB 13; Length 263;
Best Local Similarity 100.0%; Pred. No. 3.4e-125;
Matches 262; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LLAWQAFVLSNMLLAAYGSGGCFWNGHLYREDQTSAPGLRCLNWLDAQSLASAPV 60
DB 2 LLAWQAFVLSNMLLAAYGSGGCFWNGHLYREDQTSAPGLRCLNWLDAQSLASAPV 61

QY 61 SGAGNHSYCRNPDEDPGPMWCVYSGEAGVPEKPCEDLRCPETTSQALPAFTTEIOEASE 120
DB 62 SGAGNHSYCRNPDEDPGPMWCVYSGEAGVPEKPCEDLRCPETTSQALPAFTTEIOEASE 121

QY 121 GPGADEVQVFAPANALPARSEAAAVQPVIGISQVRVNSKEKKDLGTLGYVLGITMNVII 180
DB 122 GPGADEVQVFAPANALPARSEAAAVQPVIGISQVRVNSKEKKDLGTLGYVLGITMNVII 181

QY 181 IAIGAGIILGYSYKRGKDLKEQHDQKVCEREMQRIITPLSAFTNPTCEIVDEKTVVWHTS 240
DB 182 IAIGAGIILGYSYKRGKDLKEQHDQKVCEREMQRIITPLSAFTNPTCEIVDEKTVVWHTS 241

QY 241 QTPVDPQEGSTPLMGQAGTPGA 262
DB 242 QTPVDPQEGSTPLMGQAGTPGA 263

RESULT 4
US-10-210-951-44
; Sequence 44, Application US/10210951
; Publication No. US20030170228A1
; GENERAL INFORMATION:
; APPLICANT: Ashkenazi, Avi J.
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Hillan, Kenneth J.
; APPLICANT: Masters, Scot A.
; APPLICANT: Pan, James
; APPLICANT: Pitti, Robert M.
; APPLICANT: Roy, Margaret Ann
; APPLICANT: Smith, Victoria
; APPLICANT: Stone, Donna M.
; APPLICANT: Watanabe, Colin K.
; APPLICANT: Wood, William I.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE TREATMENT OF TUMOR
; FILE REFERENCE: P2931R1C1
; CURRENT APPLICATION NUMBER: US/10/210,951
; CURRENT FILING DATE: 2002-08-02
```



```

; PRIOR APPLICATION NUMBER: 60/014699
; PRIOR FILING DATE: 1996-04-01
; PRIOR APPLICATION NUMBER: 60/026943
; PRIOR FILING DATE: 1996-09-23
; PRIOR APPLICATION NUMBER: 60/059121
; PRIOR FILING DATE: 1997-07-17
; PRIOR APPLICATION NUMBER: 60/059352
; PRIOR FILING DATE: 1997-09-19
; PRIOR APPLICATION NUMBER: 60/062037
; PRIOR FILING DATE: 1997-10-10
; PRIOR APPLICATION NUMBER: 60/063755
; PRIOR FILING DATE: 1997-10-17
; PRIOR APPLICATION NUMBER: 60/063045
; PRIOR FILING DATE: 1997-10-24
; PRIOR APPLICATION NUMBER: 60/063046
; PRIOR FILING DATE: 1997-10-24
; PRIOR APPLICATION NUMBER: 60/066511
; PRIOR FILING DATE: 1997-11-24
; PRIOR APPLICATION NUMBER: 60/066772
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 258
; SEQ ID NO 44
; LENGTH: 263
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-210-951-44

```

```

Query Match      99.8%; Score 1385; DB 14; Length 263;
Best Local Similarity 99.6%; Pred. No. 6.6e-125;
Matches 261; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 LLAVQAFVLSNMLLAAYGGGCFWNGHLYREDQTSAPGLRCLNWLDAQSGLASAPV 60
DB 2 LLAVQAFVLSNMLLAAYGGGCFWNGHLYREDQTSAPGLRCLNWLDAQSGLASAPV 61

QY 61 SGAGNHSYCRNPDEDPGPPWCYVSGEAGVPEKPCEDLRCPETTSQALPAFTTEIQEASE 120
DB 62 SGAGNHSYCRNPDEDPGPPWCYVSGEAGVPEKPCEDLRCPETTSQALPAFTTEIQEASE 121

QY 121 GPGADEVQVFAPANALPARSEAAAVQPVIGISQVRVNSKEKKDLGTLGVLGITMMVII 180
DB 122 GPGADEVQVFAPANALPARSEAAAVQPVIGISQVRVNSKEKKDLGTLGVLGITMMVII 181

QY 181 IAIGAGIILGYSYKRGKDLKEHQDKVCEREMQRTITPLSAFTNPTCEIVDEKTVVHTS 240
DB 182 IAIGAGIILGYSYKRGKDLKEHQDKVCEREMQRTITPLSAFTNPTCEIVDEKTVVHTS 241

QY 241 QTPVDPQEGSTPLMQAGTPGA 262
DB 242 QTPVDPQEGSTPLMQAGTPGA 263

```

```

RESULT 5
US-10-211-884-44
; Sequence 44, Application US/10211884
; Publication No. US20030175900A1
; GENERAL INFORMATION:
; APPLICANT: Ashkenazi, Avi J.
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gutney, Austin L.
; APPLICANT: Hillan, Kenneth J.
; APPLICANT: Marsters, Scott A.
; APPLICANT: Pan, James
; APPLICANT: Pitti, Robert M.
; APPLICANT: Roy, Margaret Ann
; APPLICANT: Smith, Victoria
; APPLICANT: Stone, Donna M.
; APPLICANT: Watanabe, Colin K.
; APPLICANT: Wood, William I.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE TREATMENT OF TUMOR
; FILE REFERENCE: P2931R1C1

```

```

; CURRENT APPLICATION NUMBER: US/10/211,884
; CURRENT FILING DATE: 2002-08-02
; PRIOR APPLICATION NUMBER: 60/014699
; PRIOR FILING DATE: 1996-04-01
; PRIOR APPLICATION NUMBER: 60/026943
; PRIOR FILING DATE: 1996-09-23
; PRIOR APPLICATION NUMBER: 60/059121
; PRIOR FILING DATE: 1997-07-17
; PRIOR APPLICATION NUMBER: 60/059352
; PRIOR FILING DATE: 1997-09-19
; PRIOR APPLICATION NUMBER: 60/062037
; PRIOR FILING DATE: 1997-10-10
; PRIOR APPLICATION NUMBER: 60/063755
; PRIOR FILING DATE: 1997-10-17
; PRIOR APPLICATION NUMBER: 60/063045
; PRIOR FILING DATE: 1997-10-24
; PRIOR APPLICATION NUMBER: 60/063046
; PRIOR FILING DATE: 1997-10-24
; PRIOR APPLICATION NUMBER: 60/066511
; PRIOR FILING DATE: 1997-11-24
; PRIOR APPLICATION NUMBER: 60/066772
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 258
; SEQ ID NO 44
; LENGTH: 263
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-211-884-44

Query Match      99.8%; Score 1385; DB 14; Length 263;
Best Local Similarity 99.6%; Pred. No. 6.6e-125;
Matches 261; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 LLAVQAFVLSNMLLAAYGGGCFWNGHLYREDQTSAPGLRCLNWLDAQSGLASAPV 60
DB 2 LLAVQAFVLSNMLLAAYGGGCFWNGHLYREDQTSAPGLRCLNWLDAQSGLASAPV 61

QY 61 SGAGNHSYCRNPDEDPGPPWCYVSGEAGVPEKPCEDLRCPETTSQALPAFTTEIQEASE 120
DB 62 SGAGNHSYCRNPDEDPGPPWCYVSGEAGVPEKPCEDLRCPETTSQALPAFTTEIQEASE 121

QY 121 GPGADEVQVFAPANALPARSEAAAVQPVIGISQVRVNSKEKKDLGTLGVLGITMMVII 180
DB 122 GPGADEVQVFAPANALPARSEAAAVQPVIGISQVRVNSKEKKDLGTLGVLGITMMVII 181

QY 181 IAIGAGIILGYSYKRGKDLKEHQDKVCEREMQRTITPLSAFTNPTCEIVDEKTVVHTS 240
DB 182 IAIGAGIILGYSYKRGKDLKEHQDKVCEREMQRTITPLSAFTNPTCEIVDEKTVVHTS 241

QY 241 QTPVDPQEGSTPLMQAGTPGA 262
DB 242 QTPVDPQEGSTPLMQAGTPGA 263

```

```

RESULT 6
US-09-864-761-38457
; Sequence 38457, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharon G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; FILE REFERENCE: Aecmics-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26

```

;; PRIOR APPLICATION NUMBER: US 09/632,366  
;; PRIOR FILING DATE: 2000-08-03  
;; PRIOR APPLICATION NUMBER: GB 24263.6  
;; PRIOR FILING DATE: 2000-10-04  
;; PRIOR APPLICATION NUMBER: US 60/236,359  
;; PRIOR FILING DATE: 2000-09-27  
;; PRIOR APPLICATION NUMBER: PCT/US01/00666  
;; PRIOR FILING DATE: 2001-01-30  
;; PRIOR APPLICATION NUMBER: PCT/US01/00667  
;; PRIOR FILING DATE: 2001-01-30  
;; PRIOR APPLICATION NUMBER: PCT/US01/00664  
;; PRIOR FILING DATE: 2001-01-30  
;; PRIOR APPLICATION NUMBER: PCT/US01/00669  
;; PRIOR FILING DATE: 2001-01-30  
;; PRIOR APPLICATION NUMBER: PCT/US01/00665  
;; PRIOR FILING DATE: 2001-01-30  
;; PRIOR APPLICATION NUMBER: PCT/US01/00668  
;; PRIOR FILING DATE: 2001-01-30  
;; PRIOR APPLICATION NUMBER: PCT/US01/00663  
;; PRIOR FILING DATE: 2001-01-30  
;; PRIOR APPLICATION NUMBER: PCT/US01/00662  
;; PRIOR FILING DATE: 2001-01-30  
;; PRIOR APPLICATION NUMBER: PCT/US01/00661  
;; PRIOR FILING DATE: 2001-01-30  
;; PRIOR APPLICATION NUMBER: PCT/US01/00670  
;; PRIOR FILING DATE: 2001-01-30  
;; PRIOR APPLICATION NUMBER: US 60/234,687  
;; PRIOR FILING DATE: 2000-09-21  
;; PRIOR APPLICATION NUMBER: US 09/608,408  
;; PRIOR FILING DATE: 2000-06-30  
;; PRIOR APPLICATION NUMBER: US 09/774,203  
;; PRIOR FILING DATE: 2001-01-29  
;; NUMBER OF SEQ ID NOS: 49117  
;; SOFTWARE: Annonax Sequence Listing Engine vers. 1.1  
;; SEQ ID NO 38457  
;; LENGTH: 66  
;; TYPE: PRT  
;; ORGANISM: Homo sapiens  
;; OTHER INFORMATION: MAP TO AC002073.1  
;; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 2.2  
;; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 2.5  
;; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 2.3  
;; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 1.9  
;; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 2.2  
;; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 2.2  
;; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 2.6  
;; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 2.2  
;; OTHER INFORMATION: SWISSPROT HIT: P39881, EVALU 3.40e+00  
;; OTHER INFORMATION: EST\_HUMAN HIT: BE299406.1, EVALU 5.00e-30  
US-09-864-761-38457

Query Match 23.2%; Score 322; DB 9; Length 66;  
Best Local Similarity 100.0%; Pred. No. 2.2e-23;  
Matches 66; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 103 TTSQLPFTTEIQEASGPGADEVQVPAPANALPARSEAAVOPVIGISQVRMNSKEK 162  
DB 1 TTSQLPFTTEIQEASGPGADEVQVPAPANALPARSEAAVOPVIGISQVRMNSKEK 60  
QY 163 KDLGTL 168  
DB 61 KDLGTL 66

RESULT 7  
US-10-401-077-1  
;; Sequence 1, Application US/10401077  
;; Publication No. US20040002137A1  
;; GENERAL INFORMATION:  
;; APPLICANT: Hung, Paul Porwen  
;; APPLICANT: Wu, Bryan T. H.  
;; TITLE OF INVENTION: HUMAN TISSUE UROKINASE TYPE PLASMINOGEN

;; TITLE OF INVENTION: ACTIVATOR PRODUCTION  
;; FILE REFERENCE: 12133-006001  
;; CURRENT APPLICATION NUMBER: US/10/401.077  
;; CURRENT FILING DATE: 2003-03-27  
;; PRIOR APPLICATION NUMBER: US 60/371,013  
;; PRIOR FILING DATE: 2002-04-09  
;; NUMBER OF SEQ ID NOS: 4  
;; SOFTWARE: FastSeq for Windows Version 4.0  
;; SEQ ID NO 1  
;; LENGTH: 650  
;; TYPE: PRT  
;; ORGANISM: Homo sapiens  
US-10-401-077-1  
Query Match 11.6%; Score 160.5; DB 15; Length 650;  
Best Local Similarity 38.9%; Pred. No. 1.8e-06;  
Matches 37; Conservative 7; Mismatches 38; Indels 13; Gaps 4;  
QY 20 GSGGCFWNGHLYREDQTSFAPGLRCLNWLDAQSLASAPVS-----GAGNHSYCR 70  
DB 211 GNSDCYEDQGISYRGWTSTAESGAECTNW--NSSALAQKPSGRRPDAIRLGLGNHNYCR 268  
QY 71 NPDEDPRGWCYVSGEAGVPEKRCEDLRCPETTS 105  
DB 269 NPDERSK-PWCYVF-KAGKYSSEFCSTPACSEGNS 301  
RESULT 8  
US-09-987-457-18  
;; Sequence 18, Application US/09987457  
;; Publication No. US20030013150A1  
;; GENERAL INFORMATION:  
;; APPLICANT: Manosroi, Aranya  
;; APPLICANT: Tavapiwatana, Chatchai  
;; APPLICANT: Goetz, Friedrich  
;; APPLICANT: Werner, Rolf-Guenther  
;; TITLE OF INVENTION: Methods for Large Scale Protein Production in Prokaryotes  
;; FILE REFERENCE: 0652.2180001  
;; CURRENT APPLICATION NUMBER: US/09/987.457  
;; CURRENT FILING DATE: 2001-11-14  
;; PRIOR APPLICATION NUMBER: 60/268,573  
;; PRIOR FILING DATE: 2001-02-15  
;; PRIOR APPLICATION NUMBER: GB 00 27 782.2  
;; PRIOR FILING DATE: 2000-11-14  
;; NUMBER OF SEQ ID NOS: 18  
;; SOFTWARE: PatentIn Ver. 2.1  
;; SEQ ID NO 18  
;; LENGTH: 527  
;; TYPE: PRT  
;; ORGANISM: Homo sapiens (CPA)  
US-09-987-457-18  
Query Match 11.1%; Score 154.5; DB 10; Length 527;  
Best Local Similarity 39.6%; Pred. No. 5.2e-06;  
Matches 36; Conservative 6; Mismatches 36; Indels 13; Gaps 4;  
QY 24 CFWDNGHLYREDQTSFAPGLRCLNWLDAQSLASAPVS-----GAGNHSYCRNPDE 74  
DB 92 CYEDQGISYRGWTSTAESGAECTNW--NSSALAQKPSGRRPDAIRLGLGNHNYCRNDR 149  
QY 75 DPRGWCYVSGEAGVPEKRCEDLRCPETTS 105  
DB 150 DSK-PWCYVF-KAGKYSSEFCSTPACSEGNS 178  
RESULT 9  
US-09-987-455-19  
;; Sequence 19, Application US/09987455  
;; Publication No. US20030049729A1  
;; GENERAL INFORMATION:  
;; APPLICANT: Aranya Manosroi  
;; APPLICANT: Jiradej Manosroi



```

; APPLICANT: Zopf, David
; APPLICANT: Bayer, Robert
; APPLICANT: Hakes, David
; APPLICANT: Chen, Xi
; APPLICANT: Bove, Caryn
; TITLE OF INVENTION: ALPHA GALACTOSIDASE A: REMODELING AND GLYCOCONJUGATION OF ALPHA
; TITLE OF INVENTION: GALACTOSIDASE A
; FILE REFERENCE: 040853-01-5082
; CURRENT APPLICATION NUMBER: US/10/411,037
; CURRENT FILING DATE: 2003-04-09
; PRIOR APPLICATION NUMBER: US 60/328,523
; PRIOR FILING DATE: 2001-10-10
; PRIOR APPLICATION NUMBER: US 60/344,692
; PRIOR FILING DATE: 2001-10-19
; PRIOR APPLICATION NUMBER: US 60/387,292
; PRIOR FILING DATE: 2002-06-07
; PRIOR APPLICATION NUMBER: US 60/391,777
; PRIOR FILING DATE: 2002-06-25
; PRIOR APPLICATION NUMBER: US 60/396,594
; PRIOR FILING DATE: 2002-07-17
; PRIOR APPLICATION NUMBER: US 60/404,249
; PRIOR FILING DATE: 2002-08-16
; PRIOR APPLICATION NUMBER: US 60/407,527
; PRIOR FILING DATE: 2002-08-28
; NUMBER OF SEQ ID NOS: 75
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 26
; LENGTH: 562
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-10-411-037-26

Query Match      11.1%; Score 154.5; DB 12; Length 562;
Best Local Similarity 39.6%; Pred. No. 5.7e-06;
Matches 36; Conservative 6; Mismatches 36; Indels 13; Gaps 4;

Qy 24 CFWDNGHLYREDOTSPAPGLRCLNWLDAQSGLASAPVS-----GAGNHSYCRNPDE 74
Db 127 CYEDQGISYRGTSWSTAESGAECTNW--NSSALAQKPYSGRRPDAILRLGLGNHNYCRNPDR 184

Qy 75 DPRGPMCVYSGEAGVPEKRCPCEDLRCPEPTS 105
Db 185 DSK-PWCYVF-KAGKYSSEFCSTPACSEGNS 213

RESULT 14
US-10-193-656-8
; Sequence 8, Application US/10193656
; Publication No. US20030096733A1
; GENERAL INFORMATION:
; APPLICANT: NY, Tor
; APPLICANT: HOLMDAHL, Rikard
; APPLICANT: Li, Jinan
; TITLE OF INVENTION: NOVEL DRUG TARGETS FOR ARTHRITIS
; FILE REFERENCE: 3810/1J577-US3
; CURRENT APPLICATION NUMBER: US/10/193,656
; CURRENT FILING DATE: 2002-07-10
; PRIOR APPLICATION NUMBER: US 60/304,461
; PRIOR FILING DATE: 2001-07-10
; PRIOR APPLICATION NUMBER: US 60/304,490
; PRIOR FILING DATE: 2001-07-10
; PRIOR APPLICATION NUMBER: US 60/305,182
; PRIOR FILING DATE: 2001-07-13
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 8
; LENGTH: 562
; TYPE: PRT
; ORGANISM: Homo sapiens
; PUBLICATION INFORMATION:
; DATABASE ACCESSION NUMBER: GenBank / P00750
; DATABASE ENTRY DATE: 1986-07-21
; RELEVANT RESIDUES: (1)..(562)

US-10-193-656-8
Query Match      11.1%; Score 154.5; DB 12; Length 562;
Best Local Similarity 39.6%; Pred. No. 5.7e-06;
Matches 36; Conservative 6; Mismatches 36; Indels 13; Gaps 4;

Qy 24 CFWDNGHLYREDOTSPAPGLRCLNWLDAQSGLASAPVS-----GAGNHSYCRNPDE 74
Db 127 CYEDQGISYRGTSWSTAESGAECTNW--NSSALAQKPYSGRRPDAILRLGLGNHNYCRNPDR 184

Qy 75 DPRGPMCVYSGEAGVPEKRCPCEDLRCPEPTS 105
Db 185 DSK-PWCYVF-KAGKYSSEFCSTPACSEGNS 213

RESULT 14
US-10-193-656-8
; Sequence 8, Application US/10193656
; Publication No. US20030096733A1
; GENERAL INFORMATION:
; APPLICANT: NY, Tor
; APPLICANT: HOLMDAHL, Rikard
; APPLICANT: Li, Jinan
; TITLE OF INVENTION: NOVEL DRUG TARGETS FOR ARTHRITIS
; FILE REFERENCE: 3810/1J577-US3
; CURRENT APPLICATION NUMBER: US/10/193,656
; CURRENT FILING DATE: 2002-07-10
; PRIOR APPLICATION NUMBER: US 60/304,461
; PRIOR FILING DATE: 2001-07-10
; PRIOR APPLICATION NUMBER: US 60/304,490
; PRIOR FILING DATE: 2001-07-10
; PRIOR APPLICATION NUMBER: US 60/305,182
; PRIOR FILING DATE: 2001-07-13
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 8
; LENGTH: 562
; TYPE: PRT
; ORGANISM: Homo sapiens
; PUBLICATION INFORMATION:
; DATABASE ACCESSION NUMBER: GenBank / P00750
; DATABASE ENTRY DATE: 1986-07-21
; RELEVANT RESIDUES: (1)..(562)

US-10-193-656-8
Query Match      11.1%; Score 154.5; DB 14; Length 562;
Best Local Similarity 39.6%; Pred. No. 5.7e-06;
Matches 36; Conservative 6; Mismatches 36; Indels 13; Gaps 4;

Qy 24 CFWDNGHLYREDOTSPAPGLRCLNWLDAQSGLASAPVS-----GAGNHSYCRNPDE 74
Db 127 CYEDQGISYRGTSWSTAESGAECTNW--NSSALAQKPYSGRRPDAILRLGLGNHNYCRNPDR 184

Qy 75 DPRGPMCVYSGEAGVPEKRCPCEDLRCPEPTS 105
Db 185 DSK-PWCYVF-KAGKYSSEFCSTPACSEGNS 213

Search completed: March 18, 2004, 13:24:44
Job time : 48.6043 secs
```

```

US-10-193-656-8
Query Match      11.1%; Score 154.5; DB 14; Length 562;
Best Local Similarity 39.6%; Pred. No. 5.7e-06;
Matches 36; Conservative 6; Mismatches 36; Indels 13; Gaps 4;

Qy 24 CFWDNGHLYREDOTSPAPGLRCLNWLDAQSGLASAPVS-----GAGNHSYCRNPDE 74
Db 127 CYEDQGISYRGTSWSTAESGAECTNW--NSSALAQKPYSGRRPDAILRLGLGNHNYCRNPDR 184

Qy 75 DPRGPMCVYSGEAGVPEKRCPCEDLRCPEPTS 105
Db 185 DSK-PWCYVF-KAGKYSSEFCSTPACSEGNS 213

RESULT 15
US-10-443-701-4
; Sequence 4, Application US/10443701
; Publication No. US20030199016A1
; GENERAL INFORMATION:
; APPLICANT: Xu, Yuan
; TITLE OF INVENTION: REVERSE-PHASE HPLC ASSAY FOR PLASMINOGEN ACTIVATORS
; FILE REFERENCE: P1788R1
; CURRENT APPLICATION NUMBER: US/10/443,701
; CURRENT FILING DATE: 2003-05-21
; PRIOR APPLICATION NUMBER: US/03/703,695
; PRIOR FILING DATE: 2000-11-01
; PRIOR APPLICATION NUMBER: US 60/163,607
; PRIOR FILING DATE: 1999-11-04
; NUMBER OF SEQ ID NOS: 4
; SEQ ID NO 4
; LENGTH: 562
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-10-443-701-4

Query Match      11.1%; Score 154.5; DB 14; Length 562;
Best Local Similarity 39.6%; Pred. No. 5.7e-06;
Matches 36; Conservative 6; Mismatches 36; Indels 13; Gaps 4;

Qy 24 CFWDNGHLYREDOTSPAPGLRCLNWLDAQSGLASAPVS-----GAGNHSYCRNPDE 74
Db 127 CYEDQGISYRGTSWSTAESGAECTNW--NSSALAQKPYSGRRPDAILRLGLGNHNYCRNPDR 184

Qy 75 DPRGPMCVYSGEAGVPEKRCPCEDLRCPEPTS 105
Db 185 DSK-PWCYVF-KAGKYSSEFCSTPACSEGNS 213

Search completed: March 18, 2004, 13:24:44
Job time : 48.6043 secs
```

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

QM protein - protein search, using sw model

Run on: March 17, 2004, 07:00:39 ; Search time 19.3943 Seconds  
697.420 Million cell updates/sec

Title: US-10-057-951-2\_COPY\_2\_263  
Perfect score: 1388  
Sequence: 1 LLAWQAFVSNMLLAAYG.....PVDQEGSTPLMGQAGTGA 262

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued Patents AA:\*  
1: /cgn2\_6/ptodata/2/iaa/5A COMB pep.\*  
2: /cgn2\_6/ptodata/2/iaa/5B COMB pep.\*  
3: /cgn2\_6/ptodata/2/iaa/6A COMB pep.\*  
4: /cgn2\_6/ptodata/2/iaa/6B COMB pep.\*  
5: /cgn2\_6/ptodata/2/iaa/PTCT COMB pep.\*  
6: /cgn2\_6/ptodata/2/iaa/backfiles1.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1388	100.0	263	4	US-09-411-977-2
2	163.5	11.8	472	2	US-08-811-949-63
3	154.5	11.1	437	2	US-08-811-949-49
4	154.5	11.1	437	2	US-08-811-949-51
5	154.5	11.1	437	2	US-08-811-949-55
6	154.5	11.1	437	2	US-08-811-949-57
7	154.5	11.1	527	1	US-07-609-5108-16
8	154.5	11.1	527	2	US-08-811-949-39
9	154.5	11.1	527	5	PCT-US91-01025A-2
10	154.5	11.1	527	6	5185259-8
11	154.5	11.1	527	6	5205913-1
12	154.5	11.1	546	6	5200340-6
13	154.5	11.1	562	2	US-08-811-949-43
14	154.5	11.1	562	2	US-08-560-038A-50
15	154.5	11.1	562	2	US-08-883-795A-38
16	154.5	11.1	562	4	US-09-703-695A-4
17	154.5	11.1	562	6	5185259-3
18	154.5	11.1	562	6	5200340-2
19	154.5	11.1	562	6	5344773-2
20	148.5	10.7	83	2	US-08-811-949-2
21	146.5	10.6	655	1	US-08-148-910-12
22	146.5	10.6	655	1	US-08-448-937A-12
23	142	10.2	160	2	US-08-612-788-35
24	142	10.2	160	3	US-09-066-028-35
25	142	10.2	160	4	US-09-335-325-35
26	142	10.2	250	2	US-08-612-788-30
27	142	10.2	250	3	US-09-066-028-30

28	142	10.2	250	4	US-09-335-325-30	Sequence 30, Appl
29	142	10.2	339	1	US-08-248-629A-3	Sequence 3, Appl
30	142	10.2	339	1	US-08-451-932-3	Sequence 3, Appl
31	142	10.2	339	1	US-08-452-260-3	Sequence 3, Appl
32	142	10.2	339	1	US-08-326-785-3	Sequence 3, Appl
33	142	10.2	339	2	US-08-612-788-3	Sequence 3, Appl
34	142	10.2	339	2	US-08-605-598B-3	Sequence 3, Appl
35	142	10.2	339	2	US-08-429-743-3	Sequence 3, Appl
36	142	10.2	339	2	US-08-866-735-3	Sequence 3, Appl
37	142	10.2	339	3	US-09-066-028-3	Sequence 3, Appl
38	142	10.2	339	4	US-09-335-325-3	Sequence 3, Appl
39	142	10.2	339	5	PCT-US95-05107-3	Sequence 3, Appl
40	142	10.2	352	2	US-08-612-788-40	Sequence 40, Appl
41	142	10.2	352	3	US-09-066-028-40	Sequence 40, Appl
42	142	10.2	352	4	US-09-335-325-40	Sequence 40, Appl
43	142	10.2	374	4	US-09-377-250-3	Sequence 2, Appl
44	142	10.2	375	4	US-09-377-250-2	Sequence 2, Appl
45	142	10.2	378	2	US-08-612-788-42	Sequence 42, Appl

ALIGNMENTS

RESULT 1  
US-09-411-977-2  
; Sequence 2, Application US/09411977  
; Patent No. 6372473  
; GENERAL INFORMATION:  
; APPLICANT: Moore, Paul A.  
; APPLICANT: Ebner, Reinhard  
; TITLE OF INVENTION: Tissue Plasminogen Activator-Like Protease  
; FILE REFERENCE: P378P1  
; CURRENT APPLICATION NUMBER: US/09/411,977  
; EARLIER FILING DATE: 1999-10-04  
; EARLIER APPLICATION NUMBER: 09/084,491  
; EARLIER FILING DATE: 1998-05-27  
; EARLIER APPLICATION NUMBER: 60/048,000  
; EARLIER FILING DATE: 1997-05-28  
; NUMBER OF SEQ ID NOS: 30  
; SOFTWARE: Patent in Ver. 2.0  
; SEQ ID NO 2  
; LENGTH: 263  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-411-977-2

Query Match	100.0%;	Score	1388;	DB	4;	Length	263;
Best Local Similarity	100.0%;	Pred. No.	6e-139;	Mismatches	0;	Indels	0;
Matches	262;	Conservative	0;	0;	0;	Gaps	0;
QY	1	LLAWQAFVSNMLLAAYGSGGCFWDNGHLYREDQTSAPGLRCLNWLDAQSGLASAPV	60				
Db	2	LLAWQAFVSNMLLAAYGSGGCFWDNGHLYREDQTSAPGLRCLNWLDAQSGLASAPV	61				
QY	61	SGAGNHSYCRNPDEDFPGFVCSGEAGVPEKRCEDLRCPEITTSQALPATTIOEASE	120				
Db	62	SGAGNHSYCRNPDEDFPGFVCSGEAGVPEKRCEDLRCPEITTSQALPATTIOEASE	121				
QY	121	GGAGDEVQVEAPANALPARSAARAAVQPVIGISQVRWNSKEKKDLGTLYVGLITMVI	180				
Db	122	GGAGDEVQVEAPANALPARSAARAAVQPVIGISQVRWNSKEKKDLGTLYVGLITMVI	181				
QY	181	TAIGAGIILGVSYKSGDKLKEQHQKVCEREMQRIITPLSAFTPTCEIVDEKTVVHTS	240				
Db	182	TAIGAGIILGVSYKSGDKLKEQHQKVCEREMQRIITPLSAFTPTCEIVDEKTVVHTS	241				
QY	241	QTPVDPQEGSTPLMGQAGTPGA	262				
Db	242	QTPVDPQEGSTPLMGQAGTPGA	263				

RESULT 2

```
US-08-811-949-63
; Sequence 63, Application US/08811949
; Patent No. 5840533
; GENERAL INFORMATION:
; APPLICANT: NIWA, MINEO
; APPLICANT: SAITO, YOSHIMASA
; APPLICANT: SASAKI, HITOSHI
; APPLICANT: HAYASHI, MASAKO
; APPLICANT: NOTANI, JOUJI
; APPLICANT: KOBAYASHI, MASAKAZU
; TITLE OF INVENTION: TISSUE PLASMINOGEN ACTIVATOR
; NUMBER OF SEQUENCES: 67
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MAIER & NEUSTADT,
; ADDRESSEE: P.C.
; STREET: 1755 S. JEFFERSON DAVIS HIGHWAY, SUITE 400
; CITY: ARLINGTON
; STATE: VA
; COUNTRY: USA
; ZIP: 22202
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/811,949
; FILING DATE: 05-MAR-1997
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: OBLON, NORMAN F.
; REGISTRATION NUMBER: 24,618
; REFERENCE/DOCKET NUMBER: 18-966-0
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 703-413-3000
; TELEFAX: 703-413-2220
; INFORMATION FOR SEQ ID NO: 63:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 472 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-811-949-63

Query Match 11.8%; Score 163.5; DB 2; Length 472;
Best Local Similarity 40.0%; Pred. No. 1.1e-08;
Matches 38; Conservative 6; Mismatches 38; Indels 13; Gaps 4;

Qy 20 GSGGCFWNGHLYREDQTSAPAGRLCLNWLDAQSLASAPVS-----GAGNHSYCR 70
Db 33 GARSCEYDQGISYRGTTAESGAECTNW--NSSALAKPYSGRRPDPRLGLGNHNYCR 90

Qy 71 NPDEPRGPWCYVSGEAGVPEKPCEDLRCPETTS 105
Db 91 NPDRLSK-PWCYVF-KAGKYSSEFCSTPACSEGNS 123

RESULT 3
US-08-811-949-49
; Sequence 49, Application US/08811949
; Patent No. 5840533
; GENERAL INFORMATION:
; APPLICANT: NIWA, MINEO
; APPLICANT: SAITO, YOSHIMASA
; APPLICANT: SASAKI, HITOSHI
; APPLICANT: HAYASHI, MASAKO
; APPLICANT: NOTANI, JOUJI
; APPLICANT: KOBAYASHI, MASAKAZU
; TITLE OF INVENTION: TISSUE PLASMINOGEN ACTIVATOR
; NUMBER OF SEQUENCES: 67
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MAIER & NEUSTADT,
; ADDRESSEE: P.C.

US-08-811-949-51
; Sequence 51, Application US/08811949
; Patent No. 5840533
; GENERAL INFORMATION:
; APPLICANT: NIWA, MINEO
; APPLICANT: SAITO, YOSHIMASA
; APPLICANT: SASAKI, HITOSHI
; APPLICANT: HAYASHI, MASAKO
; APPLICANT: NOTANI, JOUJI
; APPLICANT: KOBAYASHI, MASAKAZU
; TITLE OF INVENTION: TISSUE PLASMINOGEN ACTIVATOR
; NUMBER OF SEQUENCES: 67
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MAIER & NEUSTADT,
; ADDRESSEE: P.C.

US-08-811-949-55
; Sequence 55, Application US/08811949
; Patent No. 5840533
; GENERAL INFORMATION:
; APPLICANT: NIWA, MINEO
; APPLICANT: SAITO, YOSHIMASA
; APPLICANT: SASAKI, HITOSHI
; APPLICANT: HAYASHI, MASAKO
; APPLICANT: NOTANI, JOUJI
; APPLICANT: KOBAYASHI, MASAKAZU
; TITLE OF INVENTION: TISSUE PLASMINOGEN ACTIVATOR
; NUMBER OF SEQUENCES: 67
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MAIER & NEUSTADT,
; ADDRESSEE: P.C.

US-08-811-949-59
; Sequence 59, Application US/08811949
; Patent No. 5840533
; GENERAL INFORMATION:
; APPLICANT: NIWA, MINEO
; APPLICANT: SAITO, YOSHIMASA
; APPLICANT: SASAKI, HITOSHI
; APPLICANT: HAYASHI, MASAKO
; APPLICANT: NOTANI, JOUJI
; APPLICANT: KOBAYASHI, MASAKAZU
; TITLE OF INVENTION: TISSUE PLASMINOGEN ACTIVATOR
; NUMBER OF SEQUENCES: 67
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MAIER & NEUSTADT,
; ADDRESSEE: P.C.

US-08-811-949-63
; Sequence 63, Application US/08811949
; Patent No. 5840533
; GENERAL INFORMATION:
; APPLICANT: NIWA, MINEO
; APPLICANT: SAITO, YOSHIMASA
; APPLICANT: SASAKI, HITOSHI
; APPLICANT: HAYASHI, MASAKO
; APPLICANT: NOTANI, JOUJI
; APPLICANT: KOBAYASHI, MASAKAZU
; TITLE OF INVENTION: TISSUE PLASMINOGEN ACTIVATOR
; NUMBER OF SEQUENCES: 67
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MAIER & NEUSTADT,
; ADDRESSEE: P.C.

US-08-811-949-67
; Sequence 67, Application US/08811949
; Patent No. 5840533
; GENERAL INFORMATION:
; APPLICANT: NIWA, MINEO
; APPLICANT: SAITO, YOSHIMASA
; APPLICANT: SASAKI, HITOSHI
; APPLICANT: HAYASHI, MASAKO
; APPLICANT: NOTANI, JOUJI
; APPLICANT: KOBAYASHI, MASAKAZU
; TITLE OF INVENTION: TISSUE PLASMINOGEN ACTIVATOR
; NUMBER OF SEQUENCES: 67
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MAIER & NEUSTADT,
; ADDRESSEE: P.C.

US-08-811-949-71
; Sequence 71, Application US/08811949
; Patent No. 5840533
; GENERAL INFORMATION:
; APPLICANT: NIWA, MINEO
; APPLICANT: SAITO, YOSHIMASA
; APPLICANT: SASAKI, HITOSHI
; APPLICANT: HAYASHI, MASAKO
; APPLICANT: NOTANI, JOUJI
; APPLICANT: KOBAYASHI, MASAKAZU
; TITLE OF INVENTION: TISSUE PLASMINOGEN ACTIVATOR
; NUMBER OF SEQUENCES: 67
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MAIER & NEUSTADT,
; ADDRESSEE: P.C.

US-08-811-949-75
; Sequence 75, Application US/08811949
; Patent No. 5840533
; GENERAL INFORMATION:
; APPLICANT: NIWA, MINEO
; APPLICANT: SAITO, YOSHIMASA
; APPLICANT: SASAKI, HITOSHI
; APPLICANT: HAYASHI, MASAKO
; APPLICANT: NOTANI, JOUJI
; APPLICANT: KOBAYASHI, MASAKAZU
; TITLE OF INVENTION: TISSUE PLASMINOGEN ACTIVATOR
; NUMBER OF SEQUENCES: 67
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MAIER & NEUSTADT,
; ADDRESSEE: P.C.

US-08-811-949-79
; Sequence 79, Application US/08811949
; Patent No. 5840533
; GENERAL INFORMATION:
; APPLICANT: NIWA, MINEO
; APPLICANT: SAITO, YOSHIMASA
; APPLICANT: SASAKI, HITOSHI
; APPLICANT: HAYASHI, MASAKO
; APPLICANT: NOTANI, JOUJI
; APPLICANT: KOBAYASHI, MASAKAZU
; TITLE OF INVENTION: TISSUE PLASMINOGEN ACTIVATOR
; NUMBER OF SEQUENCES: 67
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MAIER & NEUSTADT,
; ADDRESSEE: P.C.

US-08-811-949-83
; Sequence 83, Application US/08811949
; Patent No. 5840533
; GENERAL INFORMATION:
; APPLICANT: NIWA, MINEO
; APPLICANT: SAITO, YOSHIMASA
; APPLICANT: SASAKI, HITOSHI
; APPLICANT: HAYASHI, MASAKO
; APPLICANT: NOTANI, JOUJI
; APPLICANT: KOBAYASHI, MASAKAZU
; TITLE OF INVENTION: TISSUE PLASMINOGEN ACTIVATOR
; NUMBER OF SEQUENCES: 67
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MAIER & NEUSTADT,
; ADDRESSEE: P.C.

US-08-811-949-87
; Sequence 87, Application US/08811949
; Patent No. 5840533
; GENERAL INFORMATION:
; APPLICANT: NIWA, MINEO
; APPLICANT: SAITO, YOSHIMASA
; APPLICANT: SASAKI, HITOSHI
; APPLICANT: HAYASHI, MASAKO
; APPLICANT: NOTANI, JOUJI
; APPLICANT: KOBAYASHI, MASAKAZU
; TITLE OF INVENTION: TISSUE PLASMINOGEN ACTIVATOR
; NUMBER OF SEQUENCES: 67
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MAIER & NEUSTADT,
; ADDRESSEE: P.C.

US-08-811-949-91
; Sequence 91, Application US/08811949
; Patent No. 5840533
; GENERAL INFORMATION:
; APPLICANT: NIWA, MINEO
; APPLICANT: SAITO, YOSHIMASA
; APPLICANT: SASAKI, HITOSHI
; APPLICANT: HAYASHI, MASAKO
; APPLICANT: NOTANI, JOUJI
; APPLICANT: KOBAYASHI, MASAKAZU
; TITLE OF INVENTION: TISSUE PLASMINOGEN ACTIVATOR
; NUMBER OF SEQUENCES: 67
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MAIER & NEUSTADT,
; ADDRESSEE: P.C.

US-08-811-949-95
; Sequence 95, Application US/08811949
; Patent No. 5840533
; GENERAL INFORMATION:
; APPLICANT: NIWA, MINEO
; APPLICANT: SAITO, YOSHIMASA
; APPLICANT: SASAKI, HITOSHI
; APPLICANT: HAYASHI, MASAKO
; APPLICANT: NOTANI, JOUJI
; APPLICANT: KOBAYASHI, MASAKAZU
; TITLE OF INVENTION: TISSUE PLASMINOGEN ACTIVATOR
; NUMBER OF SEQUENCES: 67
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MAIER & NEUSTADT,
; ADDRESSEE: P.C.

US-08-811-949-99
; Sequence 99, Application US/08811949
; Patent No. 5840533
; GENERAL INFORMATION:
; APPLICANT: NIWA, MINEO
; APPLICANT: SAITO, YOSHIMASA
; APPLICANT: SASAKI, HITOSHI
; APPLICANT: HAYASHI, MASAKO
; APPLICANT: NOTANI, JOUJI
; APPLICANT: KOBAYASHI, MASAKAZU
; TITLE OF INVENTION: TISSUE PLASMINOGEN ACTIVATOR
; NUMBER OF SEQUENCES: 67
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MAIER & NEUSTADT,
; ADDRESSEE: P.C.
```

```
; NAME: OBLON, NORMAN F.
; REGISTRATION NUMBER: 24,618
; REFERENCE/DOCKET NUMBER: 18-966-0
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 703-413-3000
; TELEFAX: 703-413-2220
; INFORMATION FOR SEQ ID NO: 51:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 437 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-811-949-51

Query Match      11.1%; Score 154.5; DB 2; Length 437;
Best Local Similarity 39.6%; Pred. No. 8.5e-08;
Matches 36; Conservative 6; Mismatches 36; Indels 13; Gaps 4;

QY 24 CFWDNGHLYREDQTSAPGLRCLNWLDAQSLASAPVS-----GAGNHSYCRNPDE 74
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 2 CYEDQGISYRGWTSTAESGAECTNW--NSSALAQKPYSGRRPDAILRLGLGNHNYCRNPDR 59
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 75 DPRGPMCVYSGEAGVPEKRPCEDLRCPETTS 105
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 60 DSK-PWCYVF-KAGKYSSEFCSTPACSEGNS 88
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

RESULT 5
US-08-811-949-55
; Sequence 55, Application US/08811949
; Patent No. 5840533
; GENERAL INFORMATION:
; APPLICANT: NIWA, MINEO
; APPLICANT: SAITO, YOSHIMASA
; APPLICANT: SASAKI, HITOSHI
; APPLICANT: HAYASHI, MASAKO
; APPLICANT: NOTANI, JOUJI
; APPLICANT: KOBAYASHI, MASAKAZU
; TITLE OF INVENTION: TISSUE PLASMINOGEN ACTIVATOR
; NUMBER OF SEQUENCES: 67
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: OBLON, SPIVAK, MCLELLAND, MAIER & NEUSTADT,
; ADDRESS: P.C.
; STREET: 1755 S. JEFFERSON DAVIS HIGHWAY, SUITE 400
; CITY: ARLINGTON
; STATE: VA
; COUNTRY: USA
; ZIP: 22202
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA: US/08/811,949
; FILING DATE: 05-MAR-1997
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: OBLON, NORMAN F.
; REGISTRATION NUMBER: 24,618
; REFERENCE/DOCKET NUMBER: 18-966-0
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 703-413-3000
; TELEFAX: 703-413-2220
; INFORMATION FOR SEQ ID NO: 57:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 437 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-811-949-57

Query Match      11.1%; Score 154.5; DB 2; Length 437;
Best Local Similarity 39.6%; Pred. No. 8.5e-08;
Matches 36; Conservative 6; Mismatches 36; Indels 13; Gaps 4;

QY 24 CFWDNGHLYREDQTSAPGLRCLNWLDAQSLASAPVS-----GAGNHSYCRNPDE 74
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 2 CYEDQGISYRGWTSTAESGAECTNW--NSSALAQKPYSGRRPDAILRLGLGNHNYCRNPDR 59
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 75 DPRGPMCVYSGEAGVPEKRPCEDLRCPETTS 105
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 60 DSK-PWCYVF-KAGKYSSEFCSTPACSEGNS 88
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

RESULT 6
US-08-811-949-57
; Sequence 57, Application US/08811949
; Patent No. 5840533
; GENERAL INFORMATION:
; APPLICANT: NIWA, MINEO
; APPLICANT: SAITO, YOSHIMASA
; APPLICANT: SASAKI, HITOSHI
; APPLICANT: HAYASHI, MASAKO
; APPLICANT: NOTANI, JOUJI
; APPLICANT: KOBAYASHI, MASAKAZU
; TITLE OF INVENTION: TISSUE PLASMINOGEN ACTIVATOR
; NUMBER OF SEQUENCES: 67
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: OBLON, SPIVAK, MCLELLAND, MAIER & NEUSTADT,
; ADDRESS: P.C.
; STREET: 1755 S. JEFFERSON DAVIS HIGHWAY, SUITE 400
; CITY: ARLINGTON
; STATE: VA
; COUNTRY: USA
; ZIP: 22202
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA: US/08/811,949
; FILING DATE: 05-MAR-1997
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: OBLON, NORMAN F.
; REGISTRATION NUMBER: 24,618
; REFERENCE/DOCKET NUMBER: 18-966-0
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 703-413-3000
; TELEFAX: 703-413-2220
; INFORMATION FOR SEQ ID NO: 57:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 437 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-811-949-57

Query Match      11.1%; Score 154.5; DB 2; Length 437;
Best Local Similarity 39.6%; Pred. No. 8.5e-08;
Matches 36; Conservative 6; Mismatches 36; Indels 13; Gaps 4;

QY 24 CFWDNGHLYREDQTSAPGLRCLNWLDAQSLASAPVS-----GAGNHSYCRNPDE 74
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 2 CYEDQGISYRGWTSTAESGAECTNW--NSSALAQKPYSGRRPDAILRLGLGNHNYCRNPDR 59
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 75 DPRGPMCVYSGEAGVPEKRPCEDLRCPETTS 105
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 60 DSK-PWCYVF-KAGKYSSEFCSTPACSEGNS 88
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

RESULT 7
US-07-609-510B-16
; Sequence 16, Application US/07609510B
```

```

;
; ATTORNEY/AGENT INFORMATION:
; NAME: OBLON, NORMAN F.
; REGISTRATION NUMBER: 24,618
; REFERENCE/DOCKET NUMBER: 18-966-0
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 703-413-3000
; TELEFAX: 703-413-2220
; INFORMATION FOR SEQ ID NO: 39:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 527 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-811-949-39
;
; Query Match 11.1%; Score 154.5; DB 2; Length 527;
; Best Local Similarity 39.6%; Pred. No. 1.1e-07;
; Matches 36; Conservative 6; Mismatches 36; Indels 13; Gaps 4;
;
QY 24 CFWDNGHLTREDQTSPPAGRLCLNWLDAOSGLASAPVS-----GAGNHSYCRNPDE 74
Db 92 CYEDQGISYRGVWSTAESGAETNW--NSSALAQKPYSGRRPDAILRLGLGNHNYCRNPDR 149
QY 75 DRPGPCWYSGEAGVPEKPCEDLRCPETTS 105
Db 150 DSK-PCYVP-KAGKYSSEFCSTPACSEGS 178
;
; RESULT 9
; PCT-US91-01025A-2
; Sequence 2, Application PC/TUS9101025A
; GENERAL INFORMATION:
; APPLICANT: Genentech, Inc.
; TITLE OF INVENTION: Tissue Plasminogen Activator Having Fibrin
; TITLE OF INVENTION: Specific Properties
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genentech, Inc.
; STREET: 460 Point San Bruno Blvd
; CITY: South San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94020
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 5.25 inch, 360 Kb floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: patin (Genentech)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US91/01025A
; FILING DATE: 19910214
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/486,657
; FILING DATE: 1 March 1990
; ATTORNEY/AGENT INFORMATION:
; NAME: Hasak, Janet E.
; REGISTRATION NUMBER: 28,616
; REFERENCE/DOCKET NUMBER: 454P2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415/266-1896
; TELEFAX: 415/952-9881
; TELE: 910/371-7168
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 527 amino acids
; TYPE: AMINO ACID
; TOPOLOGY: linear
; PCT-US91-01025A-2
;
; Query Match 11.1%; Score 154.5; DB 5; Length 527;
; Best Local Similarity 39.6%; Pred. No. 1.1e-07;
; Matches 36; Conservative 6; Mismatches 36; Indels 13; Gaps 4;

```





```

; TELEFAX: 703-413-2220
; INFORMATION FOR SEQ ID NO: 43:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 562 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-811-949-43

Query Match 11.1%; Score 154.5; DB 2; Length 562;
Best Local Similarity 39.6%; Pred. No. 1.2e-07;
Matches 36; Conservative 6; Mismatches 36; Indels 13; Gaps 4

QY 24 CFWDNGHLVREDQTSAPGLRCLANWLDAQSLASAPVS-----GAGNHSYCRNPDE 74
DB 127 CYEDQGISVGTWSTAEGACTNW--NSSLAQKPSGRRPDAIRLGLGNHNYCENPDR 184
QY 75 DPRGPMCVYSGEAGVPEKRPCEDLRCPETTS 105
DB 185 DSK-PMCVYVF-KAGKYSSEFCSTPACSEGS 213

RESULT 14
US-08-560-098A-50
; Sequence 50, Application US/08560098A
; Patent No. 5976841
; GENERAL INFORMATION:
; APPLICANT: WRENDT, Stephan
; APPLICANT: HEINZEL-WIELAND, Regina
; TITLE OF INVENTION: STEFFENS, Gerd Josef
; TITLE OF INVENTION: Proteins having Fibrinolytic and
; TITLE OF INVENTION: Coagulation-inhibiting Properties
; NUMBER OF SEQUENCES: 60
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Evenson, McKeown, Edwards & Lenahan
; STREET: 1200 G Street, N.W., Suite 700
; CITY: Washington
; STATE: DC
; COUNTRY: USA
; ZIP: 20005
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/560,098A
; FILING DATE: 17-NOV-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: P 44 40 892.7
; FILING DATE: 17-NOV-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: EVANS, Joseph D.
; REGISTRATION NUMBER: 26,269
; REFERENCE/DOCKET NUMBER: 148/42448
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 628-8800
; TELEFAX: (202) 628-8844
; INFORMATION FOR SEQ ID NO: 50:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 562 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-560-098A-50

```

```

Query Match      11.1%; Score 154.5; DB 2; Length 562;
Best Local Similarity 39.6%; Pred. No. 1.2e-07;
Matches 36; Conservative 6; Mismatches 36; Indels 13; Gaps 4;

Qy      24  CFWDNGHLYREDQTSAPGLRCLNWLDAQGLASAPVS-----GAGNHSYCRNPDE 74
      |:      ::

```

```

Db      127 CYEQGISYRGTWSTAESGAECTNW--NSSALAQKPYSGRRRPPDAIRLGLGNHNYCRNPDR 184
Qy      75 DPRGPWCYVSGEAGVPKEKPCEDLRCPETTS 105
       : : ||||| : || |
Db      185 DSK-PWCYVF-KAGKYSSSEFCSTPACSEGNS 213
       : : ||||| : || |

RESULT 15
US-08-883-795A-38
; Sequence 38, Application US/08883795A
; Patent No. 5985607
; GENERAL INFORMATION:
; APPLICANT: Delcuve, Genevieve
; TITLE OF INVENTION: Recombinant DNA Molecules and Expression
; TITLE OF INVENTION: Vectors for Tissue Plasminogen Activator
; NUMBER OF SEQUENCES: 39
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: BERESKIN & PARR
; STREET: 40 King Street West
; CITY: Toronto
; STATE: Ontario
; COUNTRY: Canada
; ZIP: M5H 3Y2
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/883,795A
; FILING DATE: 27-JUN-1997
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Gravelle, Micheline
; REGISTRATION NUMBER: 40,261
; REFERENCE/DOCKET NUMBER: 7841-062
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (416) 364-7311
; TELEFAX: (416) 361-1398
; INFORMATION FOR SEQ ID NO: 38:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 562 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein (-PA)
US-08-883-795A-38

Query March 11.1%; Score 154.5; DB 2; Length 562;
Best Local Similarity 39.6%; Pred. No. 1.2e-07;
Matches 36; Conservative 6; Mismatches 36; Indels 13; Gaps 4

Qy      24 CFWNGHLHYREDQTSAPGLEKCLNWLDAQSLASAPVS-----CAGNHSYCRNPDE 74
       : : ||||| : || |
Db      127 CYEQGISYRGTWSTAESGAECTNW--NSSALAQKPYSGRRRPPDAIRLGLGNHNYCRNPDR 184
       : : ||||| : || |
Qy      75 DPRGPWCYVSGEAGVPKEKPCEDLRCPETTS 105
       : : ||||| : || |
Db      185 DSK-PWCYVF-KAGKYSSSEFCSTPACSEGNS 213
       : : ||||| : || |

Search completed: March 17, 2004, 07:09:06
Job time : 20.3943 secs

```

Search completed: March 17, 2004, 07:09:06  
Job time : 20.3943 secs

GenCore version 5.1.6  
Copyright (C) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 17, 2004, 06:59:28 ; Search time 16.9367 Seconds  
(without alignments)  
1374.429 Million cell updates/sec

Title: US-10-057-951-2\_COPY\_22\_263

Perfect score: 1289  
Sequence: 1 SGGCFWNGHLYREDQTPA.....PVDPOEGSTPLMGQAGTGA 242

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283366 segs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :  
1: PIR 78:\*  
2: PIR1:\*  
3: PIR2:\*  
4: PIR3:\*  
4: PIR4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	154.5	12.0	562	1 UKHUT	t-plasminogen acti
2	154	11.9	559	1 A35029	t-plasminogen acti
3	147.5	11.4	559	1 A29941	t-plasminogen acti
4	146.5	11.4	655	1 A46688	hepatocyte growth
5	145.5	11.3	291	2 I38098	t-plasminogen acti
6	142	11.0	810	1 PLRU	plasmin (EC 3.4.21
7	138	10.7	169	2 A40522	plasmin (EC 3.4.21
8	138	10.7	433	1 UK8AY	u-plasminogen acti
9	135.5	10.5	442	1 UKPG	u-plasminogen acti
10	134	10.4	431	2 JS0599	t-plasminogen acti
11	134	10.4	477	1 A34369	t-plasminogen acti
12	134	10.4	477	2 JS0598	t-plasminogen acti
13	133.5	10.4	716	1 JCS061	macrophage-stimula
14	132	10.2	431	1 UKHU	u-plasminogen acti
15	130.5	10.1	716	1 A40332	macrophage-stimula
16	129	10.0	810	2 I46260	plasmin (EC 3.4.21
17	127	9.9	394	2 JS0600	t-plasminogen acti
18	127	9.9	433	1 JN0560	u-plasminogen acti
19	126.5	9.8	728	1 A60185	hepatocyte growth
20	124.5	9.7	810	2 B30848	plasmin (EC 3.4.21
21	124	9.6	622	1 TBHU	thrombin (EC 3.4.2
22	124	9.6	728	1 JH0579	hepatocyte growth
23	124	9.6	812	1 PLBO	plasmin (EC 3.4.21
24	123.5	9.6	434	1 A35005	u-plasminogen acti
25	123	9.5	593	2 S45281	coagulation factor
26	122.5	9.5	4548	1 S00657	apoptosein(a) (EC
27	122	9.5	603	2 S28941	coagulation factor
28	121.5	9.4	728	1 A35644	hepatocyte growth
29	120	9.3	790	1 PLPG	plasmin (EC 3.4.21

RESULT 1  
UKHUT  
t-Plasminogen activator (EC 3.4.21.68) precursor [validated] - human  
N:Alternate names: t-PA, tissue plasminogen activator  
C:Species: Homo sapiens (man)  
C:Date: 14-Nov-1983 #sequence, revision 14-Nov-1983 #text, change 08-Dec-2000  
C:Accession: A94004; A23529; J70562; A93293; S02125; A91343; A93951; A91322; A54645; I6  
R:NY, T.; Elgh, F.; Lund, B.  
Proc. Natl. Acad. Sci. U.S.A. 81, 5355-5359, 1984  
A:Title: The structure of the human tissue-type plasminogen activator gene: correlation  
A:Reference number: A94004; MUID:84298137; PMID:6089198  
A:Accession: A94004  
A:Molecule type: DNA  
A:Residues: 1-562 <NYT>  
A:Cross-references: GB:L00141  
A:Note: the codon given for residue 93 (ACC) is inconsistent with the authors' translat  
R:Friezner Degen, S.J.; Rajput, B.; Reich, E.  
J. Biol. Chem. 261, 6972-6985, 1986  
A:Title: The human tissue plasminogen activator gene.  
A:Reference number: A23529; MUID:86196143; PMID:3009482  
A:Accession: A23529  
A:Molecule type: DNA  
A:Residues: 1-562 <DSG>  
A:Cross-references: GB:X03021; NID:g339817; PIDN:AAA98809.1; PID:g339818  
R:Itagaki, Y.; Yasuda, H.; Morinaga, T.; Mitsuda, S.; Higashio, K.  
Agric. Biol. Chem. 55, 1225-1232, 1991  
A:Title: Purification and characterization of tissue plasminogen activator secreted by  
A:Reference number: J70562; MUID:91291340; PMID:1368681  
A:Accession: J70562  
A:Molecule type: mRNA  
A:Residues: 31-562 <ITA>  
A:Cross-references: DBJ:D01096; NID:g220128; PIDN:BAA00881.1; PID:g441174  
A:Experimental source: embryonic lung fibroblast IMR-90 cells  
A:Note: part of this sequence, including the amino end of the mature protein, was confi  
R:Pennica, D.; Holmes, W.E.; Kohr, W.J.; Harkins, R.N.; Vehar, G.A.; Ward, C.A.; Bennet  
Nature 301, 214-221, 1983  
A:Title: Cloning and expression of human tissue-type plasminogen activator cDNA in Esch  
A:Reference number: A93293; MUID:83115262; PMID:6337343  
A:Accession: A93293  
A:Molecule type: mRNA  
A:Residues: 1-562 <PEN>  
A:Cross-references: GB:L00141  
A:Experimental source: melanoma cells  
R:Sasaki, H.; Saito, Y.; Hayashi, M.; Otsuka, K.; Niwa, M.  
Nucleic Acids Res. 16, 5695, 1988  
A:Title: Nucleotide sequence of the tissue-type plasminogen activator cDNA from human f  
A:Accession: S02125; MUID:88262579; PMID:3133640  
A:Molecule type: mRNA  
A:Status: translation not shown  
A:Residues: 1-562 <SAS>  
A:Cross-references: EMBL:X07393; NID:g37243; PIDN:CAA30302.1; PID:g37244

ALIGNMENTS

A;Experimental source: fetal lung cells  
R;Kagitani, H.; Tagawa, M.; Hatanaka, K.; Ikari, T.; Saito, A.; Bando, H.; Okada, K.; Ma  
FEBS Lett. 189, 145-149, 1985  
A;Title: Expression in *Escherichia coli* of finger-domain lacking tissue-type plasminogen  
A;Reference number: A91343; MUID:85285620; PMID:3896853  
A;Accession: A91343  
A;Molecule type: mRNA  
A;Residues: 1-38, 'G', 86-433, 'E', 435-562 <KKG>  
A;Experimental source: Detroit 562 cells; ATCC 138  
R;Edlund, T.; Ny, T.; Ranby, M.; Heden, L.O.; Palm, G.; Holmgren, E.; Josephson, S.  
Proc. Natl. Acad. Sci. U.S.A. 80, 349-352, 1983  
A;Title: Isolation of cDNA sequences coding for a part of human tissue plasminogen activator  
A;Reference number: A93951; MUID:83169656; PMID:6572997  
A;Accession: A93951  
A;Molecule type: mRNA  
A;Residues: 251-358 <EDL>  
A;Experimental source: melanoma cells  
R;Pohl, G.; Kallstrom, M.; Bergsdorf, N.; Wallen, P.; Jornvall, H.  
Biochemistry 23, 3701-3707, 1984  
A;Title: Tissue plasminogen activator: peptide analyses confirm an indirectly derived amino acid  
differences.  
A;Reference number: A90488; MUID:85000468; PMID:6433976  
A;Contents: annotation; melanoma cells, partial sequence of residues 36-562, active and  
R;Pohl, G.; Kaplan, L.; Einarsson, M.; Wallen, P.; Jornvall, H.  
FEBS Lett. 166, 29-32, 1984  
A;Title: Differences between uterine and melanoma forms of tissue plasminogen activator.  
A;Reference number: A91322; MUID:84158956; PMID:6538514  
A;Accession: A91322  
A;Molecule type: protein  
A;Residues: 33-45;311-320 <POH>  
A;Experimental source: uterus  
A;Note: in the uterus, cleavage of the activation peptide may also occur after 38-Gln  
R;van Zonneveld, A.J.; Veerman, H.; Pannekoeck, H.  
J. Biol. Chem. 261, 14214-14218, 1986  
A;Reference number: A37567; MUID:87033611; PMID:3021732  
A;Contents: annotation; fibrin binding site  
R;Verheijen, J.H.; Caspers, M.P.M.; Chang, G.T.G.; de Munk, G.A.W.; Pouwels, P.H.; Engen  
EMBO J. 5, 3525-3530, 1986  
A;Title: Involvement of finger domain and kringle 2 domain of tissue-type plasminogen activator  
A;Reference number: A37568; MUID:87161761; PMID:3030730  
A;Contents: annotation; fibrin binding site  
R;Dodd, I.; Nunn, B.; Robinson, J.H.  
Thromb. Haemost. 59, 523-528, 1988  
A;Title: Isolation, identification and pharmacokinetic properties of human tissue-type plasminogen  
activator.  
A;Reference number: A69002; MUID:89044681; PMID:3142086  
A;Contents: annotation; novel forms of expressed recombinant t-PA  
R;Harris, T.J.R.; Patel, T.; Marston, F.A.O.; Little, S.; Emtage, J.S.; Opdenakker, G.;  
Mol. Biol. Med. 3, 279-292, 1986  
A;Title: Cloning of cDNA coding for human tissue-type plasminogen activator and its expression  
in mouse cells.  
A;Reference number: A54645; MUID:86284200; PMID:3050401  
A;Accession: A54645  
A;Molecule type: mRNA  
A;Residues: 1-562 <HAR>  
A;Cross-references: GB:M15518; NID:g190031; PIDN:AAA60111.1; PID:g190032  
A;Note: parts of this sequence were confirmed by peptide sequencing  
R;Reddy, V.B.; Garramone, A.J.; Sasaki, H.; Wei, C.  
DNA 7, 461-472, 1987  
A;Title: Expression of human uterine tissue-type plasminogen activator in mouse cells using  
a plasmid vector.  
A;Reference number: A60110; MUID:88054470; PMID:2824147  
A;Accession: A60110  
A;Status: translated from GB/EMBL/DBJ  
A;Molecule type: mRNA  
A;Residues: 1-562 <RES>  
A;Cross-references: GB:M18182; NID:g340176; PIDN:AAA36800.1; PID:g340177  
R;Fisher, R.; Waller, E.K.; Groesi, G.; Thompson, D.; Tizard, R.; Schleuning, W.D.  
J. Biol. Chem. 260, 11223-11230, 1985  
A;Title: Isolation and characterization of the human tissue-type plasminogen activator cDNA  
A;Reference number: I55232; MUID:85289336; PMID:3161893  
A;Accession: I55232  
A;Status: preliminary; translated from GB/EMBL/DBJ  
A;Molecule type: DNA  
A;Residues: 1-36 <RE2>  
A;Cross-references: GB:M11890; NID:g339837; PIDN:AAA61213.1; PID:g339839

C;Comment: Cleavage by plasmin or trypsin produces two chains held together by a single  
C;Comment: t-PA converts plasminogen to plasmin by hydrolyzing a single Arg-Val bond. I  
C;Comment: t-PA binds chain A of fibrin by kringle 2 and the fibronectin type I repeat.  
C;Genetics:  
A;Gene: GDB:PLAT  
A;Cross-references: GDB:119496; OMIM:173370  
A;Map position: 8p12-8p12  
A;Introns: 24/3; 39/1; 85/1; 122/1; 180/2; 211/1; 268/2; 297/1; 362/2; 408/1; 455/3; 51  
C;Superfamily: tissue plasminogen activator; EGF homology; fibronectin type I repeat ho  
C;Keywords: fibrinolysis; glycoprotein; hydrolase; kringle; plasma; serine proteinase  
F;1-23/Domain: signal sequence #status predicted <SIG>  
F;24-32/Domain: propeptide #status predicted <PRO>  
F;33-562/Product: t-plasminogen activator #status experimental <MAT>  
F;33-310/Product: t-plasminogen activator chain A #status experimental <ACH>  
F;41-78/Domain: fibronectin type I repeat homology <1F1>  
F;86-119/Domain: EGF homology <EGF>  
F;127-208/Domain: kringle homology <KR1>  
F;215-296/Domain: kringle homology <KR2>  
F;311-562/Product: t-plasminogen activator chain B #status experimental <BCH>  
F;41-71-69-78-86-97-91-108-110-119-127-208-148-190-179-203-215-296-236-278-267-291-299-  
F;152-483/Binding site: carboxydrate (Asn) (covalent) #status experimental  
F;219/Binding site: carboxydrate (Asn) (covalent) (partial) #status experimental  
F;310-311/Cleavage site: Arg-116 (plasmin, trypsin) #status experimental  
F;357-406/Active site: His, Asp #status predicted  
F;513/Active site: Ser #status experimental  
Query Match 12.0%; Score 154.5; DB 1; Length 562;  
Best Local Similarity 39.6%; Pred. No. 2.5e-05;  
Matches 36; Conservative 6; Mismatches 36; Indels 13; Gaps 4;  
QY 4 CFWNGHLYREDQTSPPAGLRCLNWLDAQSLGASAPVS-----GAGNHSYCRNPDE 54  
DB 127 CYEDQGISYRGTWSTAGSAGCTNW--NSSALACKPYSGRRPDAIRLGLGNHNYCRNPDR 184  
QY 55 DPRGFWCVSGEAGVPEKPCEDLRCPETTS 85  
DB 185 DSK-FWCYVP-KAGYSSERCFSTACSEGNS 213  
RESULT 2  
A;Accession: A35029  
t-plasminogen activator (EC 3.4.21.68) precursor - rat  
C;Species: Rattus norvegicus (Norway rat)  
C;Date: 10-Sep-1999 #sequence\_revision 10-Sep-1999 #text\_change 10-Sep-1999  
C;Accession: A35029; A31597  
R;Feng, P.; Ohlsson, M.; Ny, T.  
J. Biol. Chem. 265, 2022-2027, 1990  
A;Title: The structure of the TATA-less rat tissue-type plasminogen activator gene. Spe  
A;Reference number: A35029; MUID:90130448; PMID:2105315  
A;Accession: A35029  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-559 <FEN>  
A;Cross-references: GB:M31197; NID:g207429; PIDN:AAA42261.1; PID:g207431; GB:J05226  
R;Ny, T.; Leonardsson, G.; Hsueh, A.J.W.  
DNA 7, 671-677, 1988  
A;Title: Cloning and characterization of a cDNA for rat tissue-type plasminogen activat  
A;Reference number: A31597; MUID:89170114; PMID:3148445  
A;Accession: A31597  
A;Molecule type: mRNA  
A;Residues: 1-379, 'K', 381-559 <NYT>  
A;Cross-references: GB:M23697; NID:g530159; PIDN:AAA41812.1; PID:g530160  
C;Superfamily: tissue plasminogen activator; EGF homology; fibronectin type I repeat ho  
C;Keywords: fibrinolysis; glycoprotein; hydrolase; kringle; signal sequence #status predicted <SIG>  
F;1-17/Domain: signal sequence #status predicted <PRO>  
F;18-29/Domain: propeptide #status predicted <PRO>  
F;30-559/Product: t-plasminogen activator #status predicted <MAT>  
F;30-308/Product: t-plasminogen activator chain A #status predicted <ACH>  
F;38-75/Domain: fibronectin type I repeat homology <1F1>  
F;83-116/Domain: EGF homology <EGF>  
F;124-205/Domain: kringle homology <KR1>  
F;213-294/Domain: kringle homology <KR2>

F:309-559/Product: t-plasminogen activator chain B #status predicted <BCH>  
F:309-553/Domain: trypsin homology <TRY>  
F:308-68,66-75,83-94,88-105,107-116,124-205,145-187,176-200,213-294,234-276,265-289,297-4  
F:149,481/Binding site: carbohydrate (Asn) (covalent) #status predicted  
F:308-309/Cleavage site: Arg-Ile (plasmin, trypsin) #status predicted  
F:355,404,510/Active site: His, Asp, Ser #status predicted

Query Match 11.9%; Score 154; DB 1; Length 559;  
Best Local Similarity 32.4%; Pred. No. 2.8e-05;  
Matches 45; Conservative 14; Mismatches 54; Indels 26; Gaps 6;

QY 4 CFWDNGHLYREDQTSAPGLRCLNWLDAQSGLASAPVS-----GAGNHSYCRNPDE 54  
DB 124 CFEGQGITVGTWSTAENGAECINW--NSSALSQKPYSGARRPNAIKLGLGHNHYCRNPDR 181  
QY 55 DRPGPCVYSGAGVPEKPCEDLRCP-----TTSQALPATTIQTQSESGPG 103  
DB 182 DVK-PWCYVF-KAGKYTTBFCTPCPKGPTDCYVGKVTYRGTHSFTT--SKASCLPW 237  
QY 104 ADEVOVFAPANLPARSEA 122  
DB 238 NSMILIGKTYTAWANSQA 256

RESULT 3  
A29941  
t-plasminogen activator (EC 3.4.21.68) precursor - mouse  
C:Species: Mus musculus (house mouse)  
C:Date: 10-Sep-1999 #sequence revision 10-Sep-1999 #text\_change 10-Sep-1999  
C:Accession: A29941; S48205; S48207; S48206  
R:Rickles, R.J.; Darrow, A.L.; Strickland, S.  
J. Biol. Chem. 263, 1563-1569, 1988  
A:Title: Molecular cloning of complementary DNA to mouse tissue plasminogen activator mRNA  
A:Reference number: A29941; MUID:88087303; PMID:2826484  
A:Accession: A29941  
A:Molecule type: mRNA  
A:Residues: 1-559 <MIC>  
A:Cross-references: GB:J03520; NID:G202109; PIDN:AAA04070.1; PID:G202110  
R:rijnen, H.R.; van Heest, B.; Beelen, V.; Collen, D.  
Eur. J. Biochem. 224, 863-871, 1994  
A:Title: Characterization of the murine plasma fibrinolytic system.  
A:Reference number: S48202; MUID:95010076; PMID:7523120  
A:Accession: S48205  
A:Molecule type: protein  
A:Residues: 33-37, 'X', 39-40 <LIJ>  
A:Accession: S48207  
A:Molecule type: protein  
A:Residues: 309-316 <LI2>  
A:Accession: S48206  
A:Molecule type: protein  
A:Residues: 33-37, 'X', 39-40 <LIW>  
C:Superfamily: tissue plasminogen activator; EGF homology; fibronectin type I repeat hom  
C:Keywords: fibrinolysis; glycoprotein; hydrolase; kringle; serine proteinase  
F:1-17/Domain: signal sequence #status predicted <SIG>  
F:18-29/Domain: propeptide #status predicted <PRO>  
F:30-559/Product: t-plasminogen activator #status predicted <MAT>  
F:30-308/Product: t-plasminogen activator chain A #status predicted <ACH>  
F:38-75/Domain: fibronectin type I repeat homology <IFI>  
F:83-116/Domain: EGF homology <EGF>  
F:124-205/Domain: kringle homology <KR1>  
F:213-294/Domain: kringle homology <KR2>  
F:309-559/Product: t-plasminogen activator chain B #status predicted <BCH>  
F:309-553/Domain: trypsin homology <TRY>  
F:308-68,66-75,83-94,88-105,107-116,124-205,145-187,176-200,213-294,234-276,265-289,297-4  
F:149,481/Binding site: carbohydrate (Asn) (covalent) #status predicted  
F:308-309/Cleavage site: Arg-Ile (plasmin, trypsin) #status predicted  
F:355,404,510/Active site: His, Asp, Ser #status predicted

Query Match 11.4%; Score 147.5; DB 1; Length 559;  
Best Local Similarity 37.0%; Pred. No. 9.8e-05;  
Matches 34; Conservative 11; Mismatches 34; Indels 13; Gaps 4;

QY 4 CFWDNGHLYREDQTSAPGLRCLNWLDAQSGLASAPVS-----GAGNHSYCRNPDE 54

DB 124 CFEGQGITVGTWSTAENGAECINW--NSSVLSLPYARRPNAIKLGLGHNHYCRNPDR 181  
QY 55 DRPGPCVYSGAGVPEKPCEDLRCPETTSQ 86  
DB 182 DLK-PWCYVF-KAGKYTTBFCTPCPKGKSE 211

RESULT 4  
A46688  
hepatocyte growth factor activator (EC 3.4.21.-) precursor [validated] - human  
C:Species: Homo sapiens (man)  
C:Date: 21-Sep-1993 #sequence revision 25-Aug-1995 #text\_change 08-Dec-2000  
C:Accession: A46688  
R:Mayazawa, K.; Shimomura, T.; Kitamura, A.; Kondo, J.; Morimoto, Y.; Kitamura, N.  
J. Biol. Chem. 268, 10024-10028, 1993  
A:Title: Molecular cloning and sequence analysis of the cDNA for a human serine protease  
d coagulation factor XII.  
A:Reference number: A46688; MUID:93252878; PMID:7683665  
A:Accession: A46688  
A:Molecule type: mRNA  
A:Residues: 1-655 <MIY>  
A:Cross-references: DDBJ:D14012; NID:G219680; PIDN:BAA03113.1; PID:G219681  
A:Experimental source: liver (mRNA); serum (protein)  
A:Note: sequence extracted from NCBI backbone (NCBIN:131227, NCBI:131228)  
A:Note: parts of the sequence, including the amino ends of the heavy and light chains,  
C:Genetics:  
A:Gene: GDB:HGFAC; HGFA; HGFAP  
A:Cross-references: GDB:9954514  
A:Map position: 4p16-4p16  
C:Function:  
A:Description: activates hepatocyte growth factor by specific proteolytic cleavage  
A:Pathway: tissue repair and regeneration  
C:Superfamily: coagulation factor XII; EGF homology; fibronectin type I repeat homology  
C:Keywords: glycoprotein; hydrolase; kringle; liver; plasma; serine proteinase  
F:1-34/Domain: signal sequence #status predicted <SIG>  
F:108-148/Domain: fibronectin type II repeat homology <IF2>  
F:164-197/Domain: EGF homology <EG1>  
F:202-237/Domain: fibronectin type I repeat homology <IF1>  
F:245-278/Domain: EGF homology <EG2>  
F:286-367/Domain: kringle homology <KRG>  
F:373-407/Product: hepatocyte growth factor activator light chain #status experimental  
F:408-655/Product: hepatocyte growth factor activator heavy chain #status experimental  
F:408-641/Domain: trypsin homology <TRY>  
F:40,48,290,468,492,546/Binding site: carbohydrate (Asn) (covalent) #status predicted  
F:164-175,169-186,188-197,202-230,228-237,245-256,250-267,269-278,286-367,307-349,338-3  
F:447,497,598/Active site: His, Asp, Ser #status predicted

Query Match 11.4%; Score 146.5; DB 1; Length 655;  
Best Local Similarity 36.9%; Pred. No. 0.00014;  
Matches 41; Conservative 10; Mismatches 41; Indels 19; Gaps 6;

QY 4 CFWDNGHLYREDQTSAPGLRCLNWLDAQSGLASAPVS-----LDAQSGLASAPVSAGNHSYCRNPDE 55  
DB 286 CFLNGTGYRGVASTSAGSCLAWSNLDLYQELHVDV-VGAALLGLGPHAYCRNPND 344  
QY 56 PRGPMCVYSGAGVPEKPCEDLRCPETTSQALPATTIQTQSESGPG 101  
DB 345 ER-PWCYVVKDSALSWEYCRLEACESTL---TRVQLSPDLLATLPPEASPG 390

RESULT 5  
I38098  
t-plasminogen activator precursor, inactive endothelial splice form - human  
N:Alternate names: tissue plasminogen activator  
C:Species: Homo sapiens (man)  
C:Date: 17-May-1996 #sequence revision 17-May-1996 #text\_change 22-Jun-1999  
C:Accession: I38098; S01678  
R:Siebert, P.D.; Fong, K.  
Nucleic Acids Res. 18, 1086, 1990  
A:Title: Variant tissue type plasminogen activator (PLAT) cDNA obtained from human endo  
A:Reference number: I38098; MUID:90192128; PMID:1969145  
A:Accession: I38098

A;Status: translated from GB/EMBL/DBJ  
A;Molecule type: mRNA  
A;Residues: 1-291 <SIE>  
A;Cross-references: EMBL:X13097; NID:G35282; PIDN:CAA31489.1; PID:G35283  
C;Comment: For the main splice form, see PIR:UKHUT. This form probably does not have pro  
C;Genetics:  
A;Gene: GDB:PLAT  
A;Cross-references: GDB:119496; OMIM:173370  
A;Map position: 8p12-8p12  
A;Introns: 24/3; 39/1; 85/1; 122/1; 180/2; 211/1; 268/2  
C;Superfamily: tissue plasminogen activator; EGF homology; fibronectin type I repeat hom  
C;Keywords: alternative splicing; fibrinolysis; glycoprotein; kringle  
F;1-23/Domain: signal sequence #status predicted <SIG>  
F;24-32/Domain: propeptide #status predicted <PRO>  
F;33-291/Product: t-plasminogen activator, inactive endothelial splice form #status pred  
F;41-78/Domain: fibronectin type I repeat homology <IFA>  
F;66-119/Domain: EGF homology <EGF>  
F;127-208/Domain: kringle homology <KR1>  
F;215-291/Domain: kringle homology #status atypical <KR2>  
F;41-71,69-78,86-97,91-108,110-119,127-208,148-190,179-203/Disulfide bonds: #status pred  
Query Match 11.3%; Score 145.5; DB 2; Length 291;  
Best Local Similarity 38.5%; Pred. No. 6 9e-05;  
Matches 35; Conservative 6; Mismatches 37; Indels 13; Gaps 4;  
Qy 4 CFWDNGHLYREDQTSAPGLRCLNWLDAQSLASPV-----GAGNHSYCRNPDE 54  
Db 127 CYEDQISYRGVWSTAESGAECTNW--NSSALQAAYSGRRPDPAIRLGLGNHYCRNPDR 184  
Qy 55 DPGCPWCYVSGEAGVEKRCEDLRCPEPTS 85  
Db 185 DSK-PWCYVP-KAGYSSEFCSTPACSEGNS 213  
RESULT 6  
PLHU  
Plasmin (EC 3.4.21.7) precursor [validated] - human  
N;Alternate names: plasminogen precursor [mismomer]  
N;Contents: angiotatin; microplasmin; plasminogen  
C;Species: Homo sapiens (man)  
C;Date: 24-Apr-1984 #sequence revision 02-Dec-1994 #text change 15-Sep-2000  
C;Accession: A35229; I52242; A26646; I62738; I84609; S03735; A00929; A04627; A04625; A04  
R;Petersen, T.E.; Martzen, M.R.; Ichinose, A.; Davie, E.W.  
J. Biol. Chem. 265, 6104-6111, 1990  
A;Title: Characterization of the gene for human plasminogen, a key proenzyme in the fibr  
A;Reference number: A35229; MUID:90202879; PMID:2318848  
A;Accession: A35229  
A;Molecule type: DNA  
A;Residues: 1-810 <PET>  
A;Cross-references: GB:J05286; GB:M34276; NID:G190064; PIDN:AAA60113.1; PID:G387026  
A;Experimental source: leukocyte; lung fibroblast  
R;Malgarretti, N.; Bruno, L.; Pontoglio, M.; Candiani, G.; Meroni, G.; Ottolenghi, S.; Ta  
Biochem. Biophys. Res. Commun. 173, 1013-1018, 1990  
A;Title: Definition of the transcription initiation site of human plasminogen gene in li  
A;Reference number: I52242; MUID:91097523; PMID:2268308  
A;Accession: I52242  
A;Status: translated from GB/EMBL/DBJ  
A;Molecule type: DNA  
A;Residues: 1-16 <MAL>  
A;Cross-references: GB:M62890; NID:G190092; PIDN:AAA36454.1; PID:G553613  
R;Forsgren, M.; Raden, B.; Israelsson, M.; Larsson, K.; Heden, L.O.  
FEBS Lett. 213, 254-260, 1987  
A;Title: Molecular cloning and characterization of a full-length cDNA clone for human pl  
A;Reference number: A26646; MUID:87162490; PMID:3030813  
A;Accession: A26646  
A;Molecule type: mRNA  
A;Residues: 1-471, 'D', 473-810 <FOR>  
A;Cross-references: GB:X05199; NID:G35530; PIDN:CAA28931.1; PID:G35531  
A;Experimental source: liver  
R;Malinowski, D.P.; Sadler, J.E.; Davie, E.W.  
Biochemistry 23, 4243-4250, 1984  
A;Title: Characterization of a complementary deoxyribonucleic acid coding for human and  
A;Reference number: I45961; MUID:85023311; PMID:6148961

A;Accession: I62738  
A;Status: translated from GB/EMBL/DBJ  
A;Molecule type: mRNA  
A;Residues: 232-471, 'D', 473-810 <MAL2>  
A;Cross-references: GB:K02922; NID:G190112; PIDN:AAA60124.1; PID:G387031  
A;Accession: I84609  
A;Status: translated from GB/EMBL/DBJ  
A;Molecule type: DNA  
A;Residues: 367-419 <MAL3>  
A;Cross-references: GB:K02921; NID:G190110; PIDN:AAA60123.1; PID:G190111  
R;Brunsholz, R.A.; Lerch, P.G.; Schaller, J.; Rickli, E.E.; Lergier, W.; Manneberg, M.  
Eur. J. Biochem. 114, 465-470, 1981  
A;Title: Comparison of the primary structure of the N-terminal CNBr fragments of human,  
A;Reference number: S03735; MUID:81212097; PMID:7238497  
A;Accession: S03735  
A;Molecule type: protein  
A;Residues: 20-71, 'E', 73-76 <BRU>  
R;Sottrup-Jensen, L.; Petersen, T.E.; Magnusson, S.  
submitted to the Atlas, July 1977  
A;Reference number: A00929  
A;Accession: A00929  
A;Molecule type: protein  
A;Residues: 20-71, 'E', 73-85, 87-106, 'D', 108-360, 'E', 362-810 <SOT>  
R;Wiman, B.  
Eur. J. Biochem. 76, 129-137, 1977  
A;Title: Primary structure of the B-chain of human plasmin.  
A;Reference number: A04627; MUID:7725245; PMID:142009  
A;Accession: A04627  
A;Molecule type: protein  
A;Residues: 581-810 <W1>  
R;Wiman, B.; Wallen, P.  
Eur. J. Biochem. 50, 489-494, 1975  
A;Title: Structural relationship between "glutamic acid" and "lysine" forms of human pl  
A;Reference number: A04625; MUID:75093329; PMID:122932  
A;Accession: A04625  
A;Molecule type: protein  
A;Residues: 20-50, 'Q', 51-71, 'E', 73-85, 87-100 <W12>  
R;Wiman, B.; Wallen, P.  
Eur. J. Biochem. 59, 539-547, 1975  
A;Title: Amino-acid sequence of the cyanogen-bromide fragment from human plasminogen th  
A;Reference number: A04626; MUID:76043692; PMID:128863  
A;Accession: A04626  
A;Molecule type: protein  
A;Residues: 483-507, 'B', 509-604 <W13>  
R;Robbins, K.C.; Bernabe, P.; Arzadon, L.; Summaria, L.  
J. Biol. Chem. 248, 1631-1633, 1973  
A;Title: The primary structure of human plasminogen. II. The histidine loop of human pl  
A;Reference number: A92125; MUID:73145248; PMID:4694729  
A;Contents: annotation; active site  
R;Groskopf, W.R.; Summaria, L.; Robbins, K.C.  
J. Biol. Chem. 244, 3590-3597, 1969  
A;Title: Studies on the active center of human plasmin. Partial amino acid sequence of  
A;Reference number: A92048; MUID:69234739; PMID:4240117  
A;Contents: annotation; active site  
R;Frexler, M.; Valli, Z.; Patches, L.  
J. Biol. Chem. 257, 7401-7406, 1982  
A;Title: Structure of the omega-aminocarboxylic acid-binding sites of human plasminogen  
A;Reference number: A92382; MUID:82213905; PMID:6919539  
A;Contents: annotation; omega-aminocarboxylic acid binding sites  
R;Valli, Z.; Patches, L.  
J. Biol. Chem. 259, 13690-13694, 1984  
A;Title: The fibrin-binding site of human plasminogen. Arginines 32 and 34 are essential  
A;Reference number: A92458; MUID:85054794; PMID:6094526  
A;Contents: annotation; fibrin binding site; omega-aminocarboxylic acid binding site  
R;Cao, Y.; Ji, R.W.; Davidson, D.; Schaller, J.; Marti, D.; Soehndel, S.; McCance, S.G.  
J. Biol. Chem. 271, 29461-29467, 1996  
A;Title: Kringle domains of human angiotatin. Characterization of the anti-proliferati  
A;Reference number: A59811; MUID:97067211; PMID:8910613  
A;Contents: annotation  
R;Lijnen, H.R.; Uguw, F.; Bini, A.; Collen, D.  
Biochemistry 37, 4699-4702, 1998  
A;Title: Generation of an angiotatin-like fragment from plasminogen by stromelysin-1

A;Reference number: A58812; MUID:9548733; PMID:9548733  
 A;Contents: annotation  
 R;Tulinsky, A.; Mulichak, A.M.  
 submitted to the Brookhaven Protein Data Bank, July 1991  
 A;Reference number: A51341; PDB:1PK4  
 A;Contents: annotation; X-ray crystallography, 1.9 angstroms, residues 376-454  
 R;Tulinsky, A.; Wu, T.P.  
 submitted to the Brookhaven Protein Data Bank, July 1991  
 A;Reference number: A51488; PDB:2PK4  
 A;Contents: annotation; X-ray crystallography, 2.25 angstroms, residues 375-454  
 R;Wu, T.P.; Tulinsky, A.  
 submitted to the Brookhaven Protein Data Bank, August 1993  
 A;Reference number: A51911; PDB:1PKR  
 A;Contents: annotation; X-ray crystallography, 2.48 angstroms, residues 102-181  
 R;Padmanabhan, K.; Tulinsky, A.  
 submitted to the Brookhaven Protein Data Bank, April 1994  
 A;Reference number: A52408; PDB:1PMK  
 A;Contents: annotation; X-ray crystallography, 2.25 angstroms, residues 377-454  
 R;Tulinsky, A.; Mathews, I.I.  
 submitted to the Brookhaven Protein Data Bank, December 1995  
 A;Reference number: A65244; PDB:ICEA  
 A;Contents: annotation; X-ray crystallography, 2.1 angstroms, residues 102-181  
 R;Tulinsky, A.; Mathews, I.I.  
 submitted to the Brookhaven Protein Data Bank, December 1995  
 A;Reference number: A65245; PDB:ICEB  
 A;Contents: annotation; X-ray crystallography, 2.1 angstroms, residues 102-181  
 R;Mulichak, A.M.; Tulinsky, A.; Ravichandran, K.G.  
 Biochemistry 30, 10576-10588, 1991  
 A;Title: Crystal and molecular structure of human plasminogen kringle 4 refined at 1.9 Å  
 A;Reference number: A58819; MUID:92031502; PMID:1657148  
 A;Contents: annotation  
 R;Wu, T.P.; Padmanabhan, K.; Tulinsky, A.; Mulichak, A.M.  
 Biochemistry 30, 10589-10594, 1991  
 A;Title: The refined structure of the epsilon-aminocaproic acid complex of human plasminogen  
 A;Reference number: A58818; MUID:92031503; PMID:1657149  
 A;Contents: annotation  
 R;de Vos, A.M.; Uitsch, M.H.; Kelley, R.F.; Padmanabhan, K.; Tulinsky, A.; Westbrook, M.  
 Biochemistry 31, 270-279, 1992  
 A;Title: Crystal structure of the kringle 2 domain of tissue plasminogen activator at 2.1 Å  
 A;Reference number: A39483; MUID:92118803; PMID:1310033  
 A;Contents: annotation; X-ray crystallography, 2.4 angstroms  
 R;Stec, B.; Teeter, M.M.; Whilow, M.; Yamano, A.  
 submitted to the Brookhaven Protein Data Bank, June 1995  
 A;Reference number: A65980; PDB:1KRN  
 A;Contents: annotation; X-ray crystallography, 1.67 angstroms, residues 376-454  
 R;Rejante, M.; Llinas, M.  
 submitted to the Brookhaven Protein Data Bank, August 1996  
 A;Reference number: A65803; PDB:1HPJ  
 A;Contents: annotation; conformation by (1)H-NMR, residues 103-181  
 R;Rejante, M.; Llinas, M.  
 submitted to the Brookhaven Protein Data Bank, August 1996  
 A;Reference number: A65804; PDB:1HPK  
 A;Contents: annotation; conformation by (1)H-NMR, residues 103-181  
 R;Rejante, M.R.; Llinas, M.  
 Eur. J. Biochem. 221, 927-937, 1994  
 A;Title: (1)H-NMR assignments and secondary structure of human plasminogen kringle 1.  
 A;Reference number: S43645; MUID:94237157; PMID:8181475  
 A;Contents: annotation; conformation by (1)H-NMR  
 R;Rejante, M.R.; Llinas, M.  
 Eur. J. Biochem. 221, 939-949, 1994  
 A;Title: Solution structure of the epsilon-aminohexanoic acid complex of human plasminogen  
 A;Reference number: A58817; MUID:94237158; PMID:8181476  
 A;Contents: annotation; conformation by (1)H-NMR  
 C;Comment: Plasminogen is synthesized by the kidney and is present in plasma and many other tissues.  
 C;Comment: Plasminogen is converted to plasmin by plasminogen activators (see PIR:UKHU and PIR:FGHUGB).  
 C;Comment: Plasmin is inactivated by alpha-2-antiplasmin (see PIR:ITHUA2) immediately after release from the fibrinolytic complex.  
 C;Comment: Microplasmin is formed by autolytic cleavage of plasmin under artificial conditions.  
 C;Comment: Stromelysin 1 (see PIR:KCHUS1) acts on plasminogen to produce angiotensin. To  
 C;Genetics:  
 A;Gene: GDB:PLG

A;Cross-references: GDB:119498; OMIM:173350  
 A;Map position: 6q26-6q27  
 A;Intons: 17/1; 62/2; 98/1; 136/2; 183/1; 223/2; 263/1; 317/2; 366/1; 419/2; 480/1; 52/2  
 C;Function:  
 A;Description: dissolves the fibrin of blood clots; acts as a proteolytic factor in a variety of tissues; also activates the urokinase-type plasminogen activator  
 A;Pathway: fibrinolysis  
 C;Superfamily: plasmin; kringle homology; plasminogen-related protein precursor homolog  
 C;Keywords: angiogenesis inhibitor; blood; duplication; fibrinolysis; glycoprotein; hyaluronidase  
 F;1-96/Domain: plasminogen-related protein precursor homology <SLPH>  
 F;1-19/Domain: signal sequence #status predicted <SIG>  
 F;20-810/Product: plasminogen #status experimental <PRO>  
 F;20-96/Domain: angiotensin peptide #status experimental <APT>  
 F;79-466/Product: angiotensin #status experimental <AST>  
 F;197-580/Domain: plasmin chain A #status experimental <NAT>  
 F;197-580/Domain: plasmin chain A #status experimental <CHA>  
 F;103-181/Domain: kringle homology <KR1>  
 F;185-262/Domain: kringle homology <KR2>  
 F;275-352/Domain: kringle homology <KR3>  
 F;377-454/Domain: kringle homology <KR4>  
 F;481-560/Domain: kringle homology <KR5>  
 F;550-580,581-810/Product: microplasmin #status experimental <MMT>  
 Query Match 11.0%; Score 142; DB 1; Length 810;  
 Best Local Similarity 37.2%; Pred. No. 0.00044;  
 Matches 32; Conservative 34; Mismatches 34; Indels 12; Gaps 4;  
 QY 4 CFWDNGLHYREDQTSAPGLRCLNWLDA-----QSGLASAPVSGAGNHSYCRNPDEPRG 58  
 Db 103 CKTGNGKXKGTWSTKXGNTCKWSTSPHRFRFPAHPSEGL-EENICRPNDDPQG 161  
 QY 59 PWCYVSGEAGVPEKR--PCEDLRCP 82  
 Db 162 PWCYTTD---PEKRYDYCDILECEE 183  
 RESULT 7  
 A40522  
 plasmin (EC 3.4.21.7) precursor - rat (fragment)  
 C;Species: Rattus norvegicus (Norway rat)  
 C;Date: 28-Feb-1992 #sequence\_revision 17-Apr-1993 #text\_change 16-Jul-1999  
 C;Accession: A40522  
 R;Kanadas, J.J.; Makker, S.P.  
 J. Biol. Chem. 265, 10825-10829, 1991  
 A;Title: Identification of the rat Heymann nephritis autoantigen (GP330) as a receptor  
 A;Reference number: A40522; MUID:91250378; PMID:1645711  
 A;Accession: A40522  
 A;Status: Preliminary  
 A;Molecule type: mRNA  
 A;Residues: 1-169 <KAN>  
 A;Cross-references: GB:M62832; NID:9206215; PIDN:AAA41884.1; PID:G554488  
 A;Note: The authors translated the codon TCT for residue 76 as Ala  
 C;Superfamily: plasmin; kringle homology; plasminogen-related protein precursor homolog  
 C;Keywords: fibrinolysis; glycoprotein; hydrolase; kringle; serine proteinase  
 F;34-112/Domain: kringle homology <KRG>  
 F;34-112,55-95,83-107/Disulfide bonds: #status predicted  
 Query Match 10.7%; Score 138; DB 2; Length 169;  
 Best Local Similarity 31.8%; Pred. No. 0.00016;  
 Matches 35; Conservative 13; Mismatches 48; Indels 14; Gaps 5;  
 QY 4 CFWDNGLHYREDQTSAPGLRCLNWLDA-----QSGLASAPVSGAGNHSYCRNPDEPRG 58  
 Db 34 CYQNGKSYKGTSTTTWTKKQSVSWTSHSKTANFPDGL-EMNYCRPNDDPQG 92  
 QY 59 PWCYVSGEAGVPEKR--PCEDLRCPETTSQALPAFTTEIQEAGEGPADE 106  
 Db 93 PWCFTTD-----PSVRWEYCNLKRCSGTGGV--AESAIVPQVPSAPGTSE 136  
 RESULT 8  
 URBA  
 u-plasminogen activator (EC 3.4.21.73) precursor - yellow baboon

C:Species: Papio cynocephalus, Papio hamadryas cynocephalus (yellow baboon)  
C:Date: 31-Dec-1991 #sequence\_revision 31-Dec-1991 #text\_change 18-Jun-1999  
C:Accession: S14687; S08651  
R:Au, Y.P.T.; Wang, T.W.; Clowes, A.W.  
Nucleic Acids Res. 18, 3411, 1990  
A:Title: Nucleotide and deduced amino acid sequences of baboon urokinase-type plasminogen  
A:Reference number: S14687; MUID:90287734; PMID:2113276  
A:Accession: S14687  
A:Molecule type: mRNA  
A:Residues: 1-433 <AUY>  
A:Cross-references: EMBL:X51935; NID:G38130; PIDN:CAA36200.1; PID:G38131  
C:Superfamily: urokinase-type plasminogen activator; EGF homology; kringle homology; try  
C:Keywords: glycoprotein; heterodimer; hydrolase; kringle; serine proteinase  
F:1-20/Domain: signal sequence #status predicted <SIG>  
F:21-176/Product: plasminogen activator chain A #status predicted <ACH>  
F:30-61/Domain: EGF homology <EGF>  
F:69-150/Domain: kringle homology <KRG>  
F:178-433/Product: plasminogen activator chain B #status predicted <BCH>  
F:178-421/Domain: trypsin homology <TRY>  
F:167-298, 208-224, 216-287, 315-384, 347-363, 374-402/Disulfide bonds: #status predicted  
F:223, 274, 378/Active site: His, Asp, Ser #status predicted  
F:324/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 10.7%; Score 138; DB 1; Length 433;  
Best Local Similarity 32.7%; Pred. No. 0.00047;  
Matches 32; Conservative 14; Mismatches 36; Indels 16; Gaps 4;

QY 4 CFWDNGHLYREDOTSPAPGLRCLNWLDAQSL-----ASAPVSGAGNHSYCRNPDP 56  
DB 69 CYEENGHFYRGKATDTMGSCLANNSATVLTQTYHAHRSDALQLGLGKHNCRNPD-NR 127

QY 57 RGPWCYVSGAGVPEK-----RPECDELRCPTTSQAL 88  
DB 128 RRPWCYV--QVGLKQVQECVMHNCADGKKPSPEEL 163

RESULT 9  
UKRG  
u-plasminogen activator (EC 3.4.21.73) precursor - pig  
N:Alternate names: uPA  
C:Species: Sus scrofa domestica (domestic pig)  
C:Date: 04-Dec-1986 #sequence\_revision 17-Mar-1987 #text\_change 07-Aug-1998  
C:Accession: A00932  
R:Naganine, Y.; Pearson, D.; Altus, M.S.; Reich, E.  
Nucleic Acids Res. 12, 9525-9541, 1984  
A:Title: cDNA and gene nucleotide sequence of porcine plasminogen activator.  
A:Reference number: A00932; MUID:85087954; PMID:6096832  
A:Accession: A00932  
A:Molecule type: DNA  
A:Residues: 1-240, 'H', 242-442 <NAG1>  
A:Experimental source: kidney cell line LLC-PK1  
R:Naganine, Y.  
Submitted to the Protein Sequence Database, December 1986  
A:Reference number: A37566  
A:Contents: annotation; correction to residue 241  
C:Genetics:  
A:Introns: 19/3; 31/1; 67/1; 125/2; 165/1; 238/2; 288/1; 335/1; 384/3  
C:Superfamily: urokinase-type plasminogen activator; EGF homology; kringle homology; try  
C:Keywords: glycoprotein; heterodimer; hydrolase; kringle; serine proteinase  
F:1-20/Domain: signal sequence #status predicted <SIG>  
F:21-168/Product: urokinase-type plasminogen activator chain A #status predicted <ACH>  
F:33-64/Domain: EGF homology <EGF>  
F:72-153/Domain: kringle homology <KRG>  
F:190-430/Domain: trypsin homology <TRY>  
F:152/Binding site: carbohydrate (Asn) (covalent) #status predicted  
F:179-310, 220-236, 228-239, 324-393, 356-372, 383-411/Disulfide bonds: #status predicted  
F:235, 286, 387/Active site: His, Asp, Ser #status predicted

Query Match 10.5%; Score 135.5; DB 1; Length 442;  
Best Local Similarity 36.9%; Pred. No. 0.00078;  
Matches 38; Conservative 7; Mismatches 37; Indels 21; Gaps 5;

QY 4 CFWDNGHLYREDOTSPAPGLRCLNWLDAQSL-----ASAPVSGAGNHSYCRNPDP 56  
DB 72 CFEGNGHYSYRGKANTGTGRPCLPWNSATVLTQTYHAHRSDALQLGLGKHNCRNPD-NQ 130

QY 57 RGPWCYVSGAGVPEK-----RPECDELRCPTTSQ 86  
DB 131 RRPWCYVQVGLKQVQECVMHNCADGKKPSPEEL 173

RESULT 10  
JS0599  
t-plasminogen activator (EC 3.4.21.68) beta precursor - common vampire bat  
N:Alternate names: tissue plasminogen activator  
C:Species: Desmodus rotundus (common vampire bat)  
C:Date: 31-Mar-1992 #sequence\_revision 31-Mar-1992 #text\_change 16-Jul-1999  
C:Accession: JS0599  
R:Kraetzschmar, J.; Haendler, B.; Langer, G.; Boidol, W.; Bringmann, P.; Alagon, A.; Do  
Gene 105, 229-237, 1991  
A:Title: The plasminogen activator family from the salivary gland of the vampire bat De  
A:Reference number: JS0597; MUID:92039036; PMID:1937019  
A:Accession: JS0599  
A:Molecule type: mRNA  
A:Residues: 1-431 <KRA>  
A:Cross-references: GB:M63989; NID:G166076; PIDN:AAA31594.1; PID:G166077  
C:Superfamily: tissue plasminogen activator; EGF homology; fibronectin type I repeat ho  
C:Keywords: fibrinolysis; glycoprotein; hydrolase; kringle; serine proteinase  
F:1-21/Domain: signal sequence #status predicted <SIG>  
F:22-36/Domain: propeptide #status predicted <PRO>  
F:37-431/Product: plasminogen activator beta #status predicted <PLA>  
F:41-74/Domain: EGF homology <EGF>  
F:82-163/Domain: kringle homology <KRG>  
F:180-425/Domain: trypsin homology <TRY>  
F:41-52, 46-63, 65-74, 82-163, 103-145, 134-158, 168-299, 211-227, 219-288, 313-388/Disulfide bo  
F:139, 352/Binding site: carbohydrate (Asn) (covalent) #status predicted  
F:179-180/Cleavage site: His-Ser (plasmin) #status predicted  
F:226, 275, 382/Active site: His, Asp, Ser #status predicted  
F:345-361, 378-406/Disulfide bonds: #status predicted

Query Match 10.4%; Score 134; DB 2; Length 431;  
Best Local Similarity 38.6%; Pred. No. 0.001;  
Matches 27; Conservative 9; Mismatches 20; Indels 14; Gaps 3;

QY 4 CFWDNGHLYREDOTSPAPGLRCLNWLDAQSL-----ASAPVSGAGNHSYCRNP 53  
DB 82 CYKDGQVYRGWTSTSESGAQCINW---NSNLTRTYNGRRSDAITLGLGHNCRNPD 138

QY 54 EDRPGPWCYV 63  
DB 139 NNSK-PWCYV 147

RESULT 11  
A34369  
t-plasminogen activator (EC 3.4.21.68) precursor - false vampire bat (Megaderma lyra)  
C:Species: Megaderma lyra  
C:Date: 10-Sep-1999 #sequence\_revision 10-Sep-1999 #text\_change 10-Sep-1999  
C:Accession: A34369  
R:Gardell, S.J.; Duong, L.T.; Diehl, R.E.; York, J.D.; Hare, T.R.; Register, R.B.; Jacc  
J. Biol. Chem. 264, 17947-17952, 1989  
A:Title: Isolation, characterization, and cDNA cloning of a vampire bat salivary plasm  
A:Reference number: A34369; MUID:90036867; PMID:2509450  
A:Accession: A34369  
A:Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-477 <GAR>  
A:Cross-references: NID:G166080; PIDN:AAA31596.1; PID:G166081  
C:Superfamily: tissue plasminogen activator; EGF homology; fibronectin type I repeat ho  
C:Keywords: fibrinolysis; glycoprotein; hydrolase; kringle; serine proteinase  
F:1-21/Domain: signal sequence #status predicted <SIG>  
F:22-36/Domain: propeptide #status predicted <PRO>  
F:37-477/Product: plasminogen activator #status predicted <PLA>  
F:42-79/Domain: fibronectin type I repeat homology <1FA>  
F:87-120/Domain: EGF homology <EGF>



A;Cross-references: EMBL:X95096; NID:g1669718; PIDN:CAA64473.1; PID:g1669719  
C;Complex: disulfide-bonded heterodimer of chains derived from the same precursor  
C;Superfamily: hepatocyte growth factor; kringle homology; trypsin homology  
C;Keywords: duplication; glycoprotein; growth factor; kringle  
F;1-31/Domain: signal sequence #status predicted <SIG>  
F;32-488/Domain: macrophage-stimulating protein 1 #status predicted <MAT>  
F;32-488/Domain: macrophage-stimulating protein 1 alpha chain #status predicted <ACH>  
F;110-186/Domain: kringle homology <KR11>  
F;191-288/Domain: kringle homology <KR12>  
F;292-370/Domain: kringle homology <KR13>  
F;379-457/Domain: kringle homology <KR14>  
F;489-716/Domain: macrophage-stimulating protein 1 beta chain #status predicted <BCH>  
F;489-709/Domain: trypsin homology <TRY>  
F;72,305,620/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 10.4%; Score 133.5; DB 1; Length 716;  
Best Local Similarity 29.1%; Pred. No. 0.002; Matches 41; Conservative 9; Mismatches 48; Indels 43; Gaps 6;

QY 4 CFWDNHGHLRYEDQTSAPGLRCLNW--LDAQSGLAGAPVSGAGNHSYCRNPDEDPGPGW 60  
DB 110 CIMDNGASYRGTVARTADGLPCQAKSRFPNDHKYVTPPKNGL-EENFCRNPDGDPGRGW 168  
QY 61 CYVS-----GEAGVPEK-RPCE--DLPCEP----- 83  
DB 169 CYTNRSVRFOSGIGKSCREAVCVWCMGEDYRGEVDVTESGRCQRWDLQHPHSHPFPE 228  
QY 84 --TSQALFAFTTEIQEASEGP 102  
DB 229 KFPDKALKNDYCRNPDA SERP 249

RESULT 14  
UKHU  
N;Alternate names: cellular plasminogen activator; urokinase; urokinase-type plasminogen  
N;Contains: urokinase-type plasminogen activator chain A; urokinase-type plasminogen ac  
in form  
C;Species: Homo sapiens (man)  
C;Date: 17-Dec-1982; sequence revision 04-Dec-1986 #text change 15-Sep-2000  
C;Accession: A00931; I52209; J70102; A37561; I38100; S85783; A37562; A37563; A37564; A37565; A37566; A37567; A37568; A37569; A37570; A37571; A37572; A37573; A37574; A37575; A37576; A37577; A37578; A37579; A37580; A37581; A37582; A37583; A37584; A37585; A37586; A37587; A37588; A37589; A37590; A37591; A37592; A37593; A37594; A37595; A37596; A37597; A37598; A37599; A37600; A37601; A37602; A37603; A37604; A37605; A37606; A37607; A37608; A37609; A37610; A37611; A37612; A37613; A37614; A37615; A37616; A37617; A37618; A37619; A37620; A37621; A37622; A37623; A37624; A37625; A37626; A37627; A37628; A37629; A37630; A37631; A37632; A37633; A37634; A37635; A37636; A37637; A37638; A37639; A37640; A37641; A37642; A37643; A37644; A37645; A37646; A37647; A37648; A37649; A37650; A37651; A37652; A37653; A37654; A37655; A37656; A37657; A37658; A37659; A37660; A37661; A37662; A37663; A37664; A37665; A37666; A37667; A37668; A37669; A37670; A37671; A37672; A37673; A37674; A37675; A37676; A37677; A37678; A37679; A37680; A37681; A37682; A37683; A37684; A37685; A37686; A37687; A37688; A37689; A37690; A37691; A37692; A37693; A37694; A37695; A37696; A37697; A37698; A37699; A37700; A37701; A37702; A37703; A37704; A37705; A37706; A37707; A37708; A37709; A37710; A37711; A37712; A37713; A37714; A37715; A37716; A37717; A37718; A37719; A37720; A37721; A37722; A37723; A37724; A37725; A37726; A37727; A37728; A37729; A37730; A37731; A37732; A37733; A37734; A37735; A37736; A37737; A37738; A37739; A37740; A37741; A37742; A37743; A37744; A37745; A37746; A37747; A37748; A37749; A37750; A37751; A37752; A37753; A37754; A37755; A37756; A37757; A37758; A37759; A37760; A37761; A37762; A37763; A37764; A37765; A37766; A37767; A37768; A37769; A37770; A37771; A37772; A37773; A37774; A37775; A37776; A37777; A37778; A37779; A37780; A37781; A37782; A37783; A37784; A37785; A37786; A37787; A37788; A37789; A37790; A37791; A37792; A37793; A37794; A37795; A37796; A37797; A37798; A37799; A37800; A37801; A37802; A37803; A37804; A37805; A37806; A37807; A37808; A37809; A37810; A37811; A37812; A37813; A37814; A37815; A37816; A37817; A37818; A37819; A37820; A37821; A37822; A37823; A37824; A37825; A37826; A37827; A37828; A37829; A37830; A37831; A37832; A37833; A37834; A37835; A37836; A37837; A37838; A37839; A37840; A37841; A37842; A37843; A37844; A37845; A37846; A37847; A37848; A37849; A37850; A37851; A37852; A37853; A37854; A37855; A37856; A37857; A37858; A37859; A37860; A37861; A37862; A37863; A37864; A37865; A37866; A37867; A37868; A37869; A37870; A37871; A37872; A37873; A37874; A37875; A37876; A37877; A37878; A37879; A37880; A37881; A37882; A37883; A37884; A37885; A37886; A37887; A37888; A37889; A37890; A37891; A37892; A37893; A37894; A37895; A37896; A37897; A37898; A37899; A37900; A37901; A37902; A37903; A37904; A37905; A37906; A37907; A37908; A37909; A37910; A37911; A37912; A37913; A37914; A37915; A37916; A37917; A37918; A37919; A37920; A37921; A37922; A37923; A37924; A37925; A37926; A37927; A37928; A37929; A37930; A37931; A37932; A37933; A37934; A37935; A37936; A37937; A37938; A37939; A37940; A37941; A37942; A37943; A37944; A37945; A37946; A37947; A37948; A37949; A37950; A37951; A37952; A37953; A37954; A37955; A37956; A37957; A37958; A37959; A37960; A37961; A37962; A37963; A37964; A37965; A37966; A37967; A37968; A37969; A37970; A37971; A37972; A37973; A37974; A37975; A37976; A37977; A37978; A37979; A37980; A37981; A37982; A37983; A37984; A37985; A37986; A37987; A37988; A37989; A37990; A37991; A37992; A37993; A37994; A37995; A37996; A37997; A37998; A37999; A38000; A38001; A38002; A38003; A38004; A38005; A38006; A38007; A38008; A38009; A38010; A38011; A38012; A38013; A38014; A38015; A38016; A38017; A38018; A38019; A38020; A38021; A38022; A38023; A38024; A38025; A38

C/Comment: This enzyme is found in urine in a high molecular mass form, consisting of  
C/Comment: Urokinase-type plasminogen activator proteolytically activates plasminogen,  
C/Genetics:  
A:Gene: GDB:PLAU  
A:Cross-references: GDB:119497; OMIM:191840  
A:Map position: 10q24-10q24  
A:Introns: 19/3; 29/1; 55/1; 123/2; 154/1; 227/2; 277/1; 324/1; 373/3  
C/Function:  
A:Description: proteolytically activates plasminogen  
A:Pathway: fibrinolysis  
C:Superfamily: urokinase-type plasminogen activator; EGF homology; kringle homology; tr  
C:Keywords: fibrinolysis; glycoprotein; heterodimer; hydrolase; kringle; serine protein  
F:1-20/Domain: signal sequence #status predicted <SIG>  
F:21-43/Product: urokinase-type plasminogen activator, single chain form #status predi  
F:21-177/Product: urokinase-type plasminogen activator chain A #status experimental <MP  
F:31-62/Domain: EGF homology <EGF>  
F:70-151/Domain: kringle homology <KR>  
F:156-177/Product: urokinase-type plasminogen activator chain A1 #status experimental <  
F:179-431/Product: urokinase-type plasminogen activator chain B #status experimental <  
F:179-419/Domain: trypsin homology <TRY>  
F:31-39,33-51,53-62,70-151,91-133,122-146,168-299,209-225,217-288,313-382,345-361,372-4  
F:38/Binding site: carbohydrate (Thr) (covalent) #status predicted  
F:178-179/Cleavage site: Lys-Ile (plasma) #status experimental  
F:224,275,376/Active site: His, Asp, Ser #status experimental  
F:322/Binding site: carbohydrate (Asn) (covalent) #status experimental

Query Match 10.2%; Score 132; DB 1; Length 431;  
Best Local Similarity 32.7%; Pred. No. 0.0015;  
Matches 33; Conservative 13; Mismatches 33; Indels 22; Gaps 5;

QY 4 CFWDNHLYREDQTSAPAGRLCLNWLDA-----QSGLASAPVSGAGNHSYCRNPEDP 56  
DB 70 CYEGNGHFYRGKASTDWMGRPCLPWNSATVLQQTTHAHRSDALQLGLGKHNYCRNPD-NR 128  
QY 57 RGPNCYVSGEAGVPEKRP-----CEDLRCPETTSQAL 88  
DB 129 RRPWCYV--QVGL---KFLVQECWVHDCADGKKFSSPPEEL 164

RESULT 15

A40332  
Macrophage-stimulating protein 1 precursor - mouse  
N/Alternate names: hepatocyte growth factor-like protein  
C/Species: Mus musculus (house mouse)  
C/Date: 17-Jul-1992 #sequence\_revision 17-Jul-1992 #text\_change 18-Jun-1999  
C/Accession: A40332; B40332  
R/Degen, S.J.F.; Stuart, L.A.; Han, S.; Jamison, C.S.  
Biochemistry 30, 9781-9791, 1991  
A/Title: Characterization of the mouse cDNA and gene coding for a hepatocyte growth fac  
A/Reference number: A40332; MUID:9202017; PMID:1832957  
A/Accession: A40332  
A/Molecule type: DNA  
A/Residues: 1-716 <DEG>  
A/Cross-references: GB:M74180; NID:g193831; PIDN:AAA50166.1; PID:g193832  
A/Accession: B40332  
A/Molecule type: mRNA  
A/Residues: 1-18, P', 20-716 <DEG2>  
A/Cross-references: GB:M74181; NID:g193833; PIDN:AAA50167.1; PID:g193834  
C/Genetics:  
A/Introns: 18/1; 67/2; 105/1; 143/2; 189/1; 229/2; 269/1; 334/2; 378/1; 412/2; 458/1; 4  
C/Complex: disulfide-bonded heterodimer of chains derived from the same precursor  
C:Superfamily: hepatocyte growth factor; kringle homology; trypsin homology  
C:Keywords: duplication; glycoprotein; growth factor; kringle  
F:1-31/Domain: signal sequence #status predicted <SIG>  
F:19-488,489-716/Product: macrophage-stimulating protein 1 #status experimental <MAT>  
F:19-483/Domain: alpha chain #status experimental <ACH>  
F:110-186/Domain: kringle homology <KR1>  
F:191-268/Domain: kringle homology <KR2>  
F:292-370/Domain: kringle homology <KR3>  
F:379-457/Domain: kringle homology <KR4>  
F:484-711/Domain: beta chain #status experimental <BCH>  
F:489-709/Domain: trypsin homology <TRY>  
F:72,173,305,620/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 10.1%; Score 130.5; DB 1; Length 716;  
Best Local Similarity 30.7%; Pred. No. 0.0036;  
Matches 35; Conservative 7; Mismatches 37; Indels 35; Gaps 5;  
QY 4 CFWDNGHLYREDOTSPAGLRLNW---LDAQSLASAFVSGAGNHSYCRNPDEDPRGFW 60  
Db 110 CIMDNGSVYRGTVARTAGGLPCQAWSRRFPNDHKYTPTEKGL-BENFCRNPEDGDPGRFW 168  
QY 61 CYVS-----GEAGVPEK-RPCE--DLRCPET 83  
Db 169 CYTTRSVRFQSCGKIKTCREAVCVLCNGEDYRGVDVTEGRCORWDLQPHS 222

Search completed: March 17, 2004, 07:07:58  
Job time : 16.9367 secs

GenCore version 5.1.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 17, 2004, 06:55:43 ; Search time 9,7712 Seconds  
(without alignments)  
1289.604 Million cell updates/sec

Title: US-10-057-951-2\_COPY\_22\_263

Perfect score: 1289

Sequence: 1 SGCGFWNDGHLXREDTSPA.....PVDQEGSTPLMGQAGTPCA 242

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_42.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	156.5	12.1	566	1 TPA_BOVIN	Q28198 bos taurus
2	154.5	12.0	562	1 TPA_HUMAN	P00750 homo sapien
3	154	11.9	559	1 TPA_RAT	P19637 rattus norv
4	150	11.6	653	1 HGFA_MOUSE	Q90988 mus musculu
5	147.5	11.4	559	1 TPA_MOUSE	P11214 mus musculu
6	146.5	11.4	655	1 HGFA_HUMAN	Q04756 homo sapien
7	142	11.0	810	1 PLMN_HUMAN	P00747 homo sapien
8	138	10.7	169	1 PLMN_RAT	Q01177 rattus norv
9	138	10.7	433	1 UROK_PAPCY	P16227 papio cynoc
10	135.5	10.5	442	1 UROK_PIG	P84185 sus scrofa
11	134	10.4	431	1 URTB_DESRO	P98121 desmodus ro
12	134	10.4	477	1 URT2_DESRO	P15638 desmodus ro
13	132	10.2	431	1 UROK_HUMAN	P00749 homo sapien
14	130.5	10.1	716	1 HGFL_MOUSE	P26928 mus musculu
15	129	10.0	810	1 PLMN_ERIEU	Q29485 erinaceus e
16	127	9.9	394	1 URTG_DESRO	P49150 desmodus ro
17	127	9.9	433	1 UROK_BOVIN	Q05589 bos taurus
18	126.5	9.8	728	1 HGF_MOUSE	Q08048 mus musculu
19	124.5	9.7	810	1 PLMN_MACMU	P12545 macaca mula
20	124	9.6	622	1 THRB_HUMAN	P00734 homo sapien
21	124	9.6	728	1 HGF_HUMAN	P14210 homo sapien
22	124	9.6	812	1 PLMN_BOVIN	P06868 bos taurus
23	123.5	9.6	434	1 UROK_CHICK	P15120 gallus gall
24	123	9.5	593	1 PA12_BOVIN	P98140 bos taurus
25	122.5	9.5	4548	1 APOA_HUMAN	P08519 homo sapien
26	122	9.5	603	1 PA12_CAVPO	Q04962 cavia porce
27	121.5	9.4	728	1 HGF_RAT	P17945 rattus norv
28	120.5	9.3	333	1 PLMN_CANFA	P80009 canis fami
29	120.5	9.3	461	1 KRM2_MOUSE	Q81167 mus musculu
30	120	9.3	790	1 PLMN_PIG	P06867 sus scrofa
31	119.5	9.3	462	1 KRM2_HUMAN	Q8ncw0 homo sapien
32	119.5	9.3	477	1 URT1_DESRO	P98119 desmodus ro
33	119	9.2	433	1 UROK_MOUSE	P06869 mus musculu

34	118.5	9.2	615	1 PA12_HUMAN	P00748 homo sapien
35	118	9.2	432	1 UROK_RAT	P29598 rattus norv
36	117.5	9.1	812	1 PLMN_MOUSE	P20318 mus musculu
37	117	9.1	625	1 THRB_BOVIN	P00735 bos taurus
38	116	9.0	1420	1 APOA_MACMU	P14417 macaca mula
39	115.5	9.0	711	1 HGFL_HUMAN	P26927 homo sapien
40	114	8.8	473	1 KRM1_MOUSE	Q99n43 mus musculu
41	114	8.8	473	1 KRM1_RAT	Q92454 rattus norv
42	114	8.8	475	1 KRM1_HUMAN	Q96mu8 homo sapien
43	114	8.8	618	1 THRB_MOUSE	P19221 mus musculu
44	113.5	8.8	452	1 KRM1_XENLA	Q90y90 xenopus lae
45	110.5	8.6	685	1 ROR1_DROME	Q24488 drosophila

#### ALIGNMENTS

##### RESULT 1

TPA\_BOVIN

ID TPA\_BOVIN STANDARD; PRT; 566 AA.

AC Q28198;

DT 01-NOV-1997 (Rel. 35, Created)

DT 01-NOV-1997 (Rel. 35, Last sequence update)

DT 15-MAR-2004 (Rel. 43, Last annotation update)

DE Tissue-type plasminogen activator precursor (EC 3.4.21.68) (TPA)

DE (t-PA) (t-plasminogen activator)

GN PLAT.

OS Bos taurus (Bovine).

OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;

OC Bovidae; Bovinae; Bos.

OX NCBI\_TaxID=9913;

RN [1]

RP SEQUENCE FROM N.A.

RC TISSUE=Kidney;

RA Ravn P., Berglund L., Petersen T.E.;

RT "Cloning and characterization of the bovine plasminogen activators uPA

RT and tPA";

RL Int. Dairy J. 5:605-617(1995).

CC -!- FUNCTION: Converts the abundant, but inactive, zymogen plasminogen

CC to plasmin by hydrolyzing a single Arg-Val bond in plasminogen. By

CC controlling plasmin-mediated proteolysis, it plays an important

CC role in tissue remodeling and degradation, in cell migration and

CC many other physiological events.

CC -!- CATALYTIC ACTIVITY: Specific cleavage of Arg-|-Val bond in

CC plasminogen to form plasmin.

CC -!- SUBUNIT: Heterodimer of chain A and chain B held by a disulfide

CC bond.

CC -!- SUBCELLULAR LOCATION: Secreted; extracellular.

CC -!- PTM: THE SINGLE CHAIN, ALMOST FULLY ACTIVE ENZYME, CAN BE FURTHER

CC PROCESSED INTO A TWO-CHAIN FULLY ACTIVE FORM BY A CLEAVAGE AFTER

CC ARG-314 CATALYZED BY PLASMIN, TISSUE KALLIKREIN OR FACTOR XA.

CC -!- MISCELLANEOUS: Binds to the kringle structure of the fibrin A

CC chain. Binding to fibrin enhances its catalytic activity.

CC -!- SIMILARITY: Belongs to peptidase family S1.

CC -!- SIMILARITY: Contains 1 EGF-like domain.

CC -!- SIMILARITY: Contains 1 fibronectin type I domain.

CC -!- SIMILARITY: Contains 2 kringle domains.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----

CC EMBL; X85900; CAA59795.1; -

CC HSSP; P00750; IRTF.

CC MEROPS; S01.232; -

CC InterPro; IPR009003; Cys Ser trypsin.

CC InterPro; IPR006209; EGF-like.

CC InterPro; IPR000083; Fibrinctn.

CC DR

DR InterPro: IPR006210; IEGF.  
DR InterPro: IPR000001; Kringle.  
DR InterPro: IPR001254; Peptidase\_S1.  
DR InterPro: IPR001314; Peptidase\_S1A.  
DR Pfam: PF00008; EGF\_1.  
DR Pfam: PF00039; Inl\_1.  
DR Pfam: PF00051; kringle; 2.  
DR Pfam: PF00089; trypsin; 1.  
DR PRINTS: PR00722; CHYMOTRYPSIN.  
DR PRINTS: PR0018; KRINGLE.  
DR ProDom: PD000395; Kringle; 2.  
DR SMART: SM00181; EGF\_1.  
DR SMART: SM00058; FN1\_1.  
DR SMART: SM00130; KR; 2.  
DR SMART: SM00020; Tryp\_Spc; 1.  
DR PROSITE: PS00022; EGF\_1; 1.  
DR PROSITE: PS01186; EGF\_2; 1.  
DR PROSITE: PS00026; EGF\_3; 1.  
DR PROSITE: PS01253; FIBRINECTIN\_1; 1.  
DR PROSITE: PS00021; KRINGLE\_1; 1.  
DR PROSITE: PS00070; KRINGLE\_2; 2.  
DR PROSITE: PS00240; TRYPSIN\_DOM; 1.  
DR PROSITE: PS00134; TRYPSIN\_HIS; 1.  
DR PROSITE: PS00135; TRYPSIN\_SER; 1.  
KW Plasminogen activation; Hydrolase; Serine protease; Glycoprotein;  
KW Plasma; Kringle; EGF-like domain; Repeat; Signal.  
FT SIGNAL 1 21 BY SIMILARITY.  
FT PROPEP 22 33 TISSUE-TYPE PLASMINOGEN ACTIVATOR.  
FT CHAIN 34 566 TISSUE-TYPE PLASMINOGEN ACTIVATOR A  
FT CHAIN 34 314 CHAIN.  
FT CHAIN 315 566 TISSUE-TYPE PLASMINOGEN ACTIVATOR B  
FT CHAIN.  
FT DOMAIN 40 82 FIBRINECTIN TYPE-1.  
FT DOMAIN 83 121 EGF-LIKE.  
FT DOMAIN 128 209 KRINGLE 1.  
FT DOMAIN 219 300 KRINGLE 2.  
FT DOMAIN 315 566 SERINE PROTEASE.  
FT ACT\_SITE 361 410 CHARGE RELAY SYSTEM.  
FT ACT\_SITE 410 461 CHARGE RELAY SYSTEM.  
FT ACT\_SITE 517 517 CHARGE RELAY SYSTEM.  
FT DISULFID 42 72 BY SIMILARITY.  
FT DISULFID 70 79 BY SIMILARITY.  
FT DISULFID 87 98 BY SIMILARITY.  
FT DISULFID 92 109 BY SIMILARITY.  
FT DISULFID 111 120 BY SIMILARITY.  
FT DISULFID 128 209 BY SIMILARITY.  
FT DISULFID 149 191 BY SIMILARITY.  
FT DISULFID 160 204 BY SIMILARITY.  
FT DISULFID 219 300 BY SIMILARITY.  
FT DISULFID 240 282 BY SIMILARITY.  
FT DISULFID 271 295 BY SIMILARITY.  
FT DISULFID 303 434 INTERCHAIN (BY SIMILARITY).  
FT DISULFID 346 362 BY SIMILARITY.  
FT DISULFID 354 423 BY SIMILARITY.  
FT DISULFID 448 523 BY SIMILARITY.  
FT DISULFID 480 496 BY SIMILARITY.  
FT DISULFID 513 541 BY SIMILARITY.  
FT CARBOHYD 153 153 N-LINKED (GLCNAC... ) (POTENTIAL).  
FT CARBOHYD 487 487 N-LINKED (GLCNAC... ) (POTENTIAL).  
SQ SEQUENCE 566 AA; 63701 MW; 2EB6BBE4E32276C3 CRC64;  
Query Match 12.1%; Score 156.5; DB 1; Length 566;  
Best Local Similarity 36.7%; Pred. No. 1e-05; 39; Indels 13; Gaps 4;  
Matches 36; Conservative 10; Mismatches 10; Indels 13; Gaps 4;  
QY 4 CPWDNGHYLYREDQTSFAPGLRCLNWLDAQSLGAPVS-----GAGNHSYCRNPDE 54  
Db 128 CYKQGVAYRGVGTWSTAESGAECANW--NSSGLAMKPYSGRRPNALRLGLGNHNYCRNPDQ 185  
QY 55 DRPGWCVYSGAGVPEKRPCEDLRCPTTSQALPAFT 92  
Db 186 DSK-PWCYVF-KAGKYISEFCSTPACAKVAEEDGDCYT 221

RESULT 2  
TPA\_HUMAN STANDARD; PRT; 562 AA.  
AC P00750; O15103;  
DT 21-JUL-1986 (Rel. 01, Created)  
DT 21-JUL-1986 (Rel. 01, Last sequence update)  
DT 15-MAR-2004 (Rel. 43, Last annotation update)  
DE Tissue-type plasminogen activator precursor (EC 3.4.21.68) (tPA)  
DE (t-PA) (t-plasminogen activator) (Alteplase) (Reteplase).  
GN PLAT.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCHI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Melanoma;  
RX MEDLINE=83115262; PubMed=6337343;  
RA Pennica D., Holmes W.E., Kohr W.J., Harkins R.N., Vehar G.A.,  
RA Ward C.A., Bennett W.F., Yelverton E., Seeburg P.H., Heyneker H.L.,  
RA Goeddel D.V., Collen D.;  
RT "Cloning and expression of human tissue-type plasminogen activator  
RT cDNA in E. coli.";  
RL Nature 301:214-221(1983).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Fetal lung;  
RX MEDLINE=88262579; PubMed=3133640;  
RA Sasaki H., Saito Y., Hayashi M., Otsuka K., Niwa M.;  
RT "Nucleotide sequence of the tissue-type plasminogen activator cDNA  
RT from human fetal lung cells.";  
RL Nucleic Acids Res. 16:5695-5695(1988).  
RN [3]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=88054470; PubMed=2824147;  
RA Reddy V.B., Garramone A.J., Sasak H., Wei C.-M., Watkins P., Galli J.,  
RA Hsiung N.;  
RT "Expression of human uterine tissue-type plasminogen activator in  
RT mouse cells using BPV vectors.";  
RL DNA 6:461-472(1987).  
RN [4]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=86196143; PubMed=3009482;  
RA Friezner Degen S.J., Rajput B., Reich E.;  
RT "The human tissue plasminogen activator gene.";  
RL J. Biol. Chem. 261:6972-6985(1986).  
RN [5]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=84298137; PubMed=6089198;  
RA Ny T., Elgh F., Lund B.;  
RT "The structure of the human tissue-type plasminogen activator gene:  
RT correlation of intron and exon structures to functional and  
RT structural domains.";  
RL Proc. Natl. Acad. Sci. U.S.A. 81:5355-5359(1984).  
RN [6]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=86284200; PubMed=3090401;  
RA Harris T.J., Patel T., Marston P.A., Little S., Entage J.S.,  
RA Odenakker G., Volckaert G., Rombauts W., Billiau A., Somer P.;  
RT "Cloning of cDNA coding for human tissue-type plasminogen activator  
RT and its expression in Escherichia coli.";  
RL Mol. Biol. Med. 3:279-292(1986).  
RN [7]  
RP SEQUENCE FROM N.A. (ISOFORM SHORT).  
RC TISSUE=Umbilical vein;  
RX MEDLINE=90192129; PubMed=2107528;  
RA Siebert P.D., Fong K.;  
RT "Variant tissue-type plasminogen activator (PLAT) cDNA obtained from  
RT human endothelial cells.";  
RL Nucleic Acids Res. 18:1086-1086(1990).  
RN [8]

SEQUENCE FROM N.A.  
 RC TISSUE=Brain; PubMed=12477932;  
 RX MEDLINE=22388257; PubMed=12477932;  
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
 RA Klausner R.D., Collins P.S., Wagner L., Shenmen C.M., Schuler G.D.,  
 RA Altschul S.P., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,  
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,  
 RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,  
 RA Stapleton M., Soares M.B., Donald M.F., Casavani T.L., Schetz T.B.,  
 RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,  
 RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,  
 RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,  
 RA Villalón D.K., Muny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
 RA Fahy J., Helton E., Kettman M., Madan A., Rodrigues S., Sanchez A.,  
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,  
 RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smallos D.E.,  
 RA Scherch A., Schein J.E., Jones S.J.M., Marra M.A.,  
 RT "Generation and initial analysis of more than 15,000 full-length  
 RT human and mouse cDNA sequences.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).  
 RN [9]  
 RP SEQUENCE OF 212-361 FROM N.A.  
 RX MEDLINE=83169656; PubMed=6572897;  
 RA Edlund T., Ny T., Raanby M., Heden L.-O., Palm G., Holmgren E.,  
 RA Josephson S.,  
 RT "Isolation of cDNA sequences coding for a part of human tissue  
 RT plasminogen activator.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 80:349-352 (1983).  
 RN [10]  
 RP SEQUENCE OF 1-36 FROM N.A.  
 RX MEDLINE=85289338; PubMed=3161893;  
 RA Fisher R., Waller E.K., Grossi G., Thompson D., Tizard R.,  
 RA Schleuning W.-D.,  
 RT "Isolation and characterization of the human tissue-type plasminogen  
 RT activator structural gene including its 5' flanking region.";  
 RL J. Biol. Chem. 260:11223-11230 (1985).  
 RN [11]  
 RP SEQUENCE OF 31-562 FROM N.A.  
 RX MEDLINE=91291340; PubMed=1386861;  
 RA Itagaki Y., Yasuda H., Morinaga T., Mitsuda S., Higashio K.,  
 RT "Purification and characterization of tissue plasminogen activator  
 RT secreted by human embryonic lung diploid fibroblasts, IMR-90 cells.";  
 RL Agric. Biol. Chem. 55:1225-1232 (1991).  
 RN [12]  
 RP SEQUENCE OF 36-562.  
 RC TISSUE=Melanoma;  
 RX MEDLINE=85000468; PubMed=6433976;  
 RA Pohl G., Kallstroem M., Bergsdorf N., Wallen P., Joernvall H.,  
 RT "Tissue plasminogen activator: peptide analyses confirm an indirectly  
 RT derived amino acid sequence, identify the active site serine residue,  
 RT establish glycosylation sites, and localize variant differences.";  
 RL Biochemistry 23:3701-3707 (1984).  
 RN [13]  
 RP SEQUENCE OF 33-52 AND 311-330.  
 RC TISSUE=Melanoma;  
 RX MEDLINE=83209620; PubMed=6682760;  
 RA Wallen P., Pohl G., Bergsdorf N., Raanby M., Ny T., Joernvall H.,  
 RT "Purification and characterization of a melanoma cell plasminogen  
 RT activator.";  
 RL Eur. J. Biochem. 132:681-686 (1983).  
 RN [14]  
 RP STRUCTURE OF CARBOHYDRATES.  
 RX MEDLINE=90092112; PubMed=2513186;  
 RA Pfeiffer G., Schmidt M., Strube K.-H., Geyer R.,  
 RT "Carbohydrate structure of recombinant human uterine tissue  
 RT plasminogen activator expressed in mouse epithelial cells.";  
 RL Eur. J. Biochem. 186:273-286 (1989).  
 RN [15]  
 RP CARBOHYDRATE-LINKAGE SITE THR-96.  
 RX MEDLINE=91159408; PubMed=1900431;  
 RA Harris R.J., Leonard C.K., Guzzetta A.W., Spellman M.W.,  
 RT "Tissue plasminogen activator has an O-linked fucose attached to  
 RT threonine-61 in the epidermal growth factor domain.";  
 RL Biochemistry 30:2311-2314 (1991).  
 RN [16]  
 RP DISULFIDE BONDS IN KRINGLE 2.  
 RX MEDLINE=91244765; PubMed=1645336;  
 RA Vlahos C.J., Wilhelm O.G., Hassell T., Jaskunas S.R., Bang N.U.,  
 RT "Disulfide pairing of the recombinant kringle-2 domain of tissue  
 RT plasminogen activator produced in *Escherichia coli*.";  
 RL J. Biol. Chem. 266:10070-10072 (1991).  
 RN [17]  
 RP X-RAY CRYSTALLOGRAPHY (2.3 ÅNGSTRÖMS) OF CATALYTIC DOMAIN.  
 RX MEDLINE=96200985; PubMed=8613982;  
 RA Lamba D., Bauer M., Huber R., Fischer S., Rudolph R., Kohnert U.,  
 RA Bode W.,  
 RT "The 2.3 Å crystal structure of the catalytic domain of recombinant  
 RT two-chain human tissue-type plasminogen activator.";  
 RL J. Mol. Biol. 258:117-135 (1996).  
 RN [18]  
 RP X-RAY CRYSTALLOGRAPHY (3.1 ÅNGSTRÖMS) OF CATALYTIC DOMAIN.  
 RX MEDLINE=97449126; PubMed=9305622;  
 RA Renatus M., Engh R.A., Stubbs M.T., Huber R., Fischer S., Kohnert U.,  
 RA Bode W.,  
 RT "Lysine 156 promotes the anomalous proenzyme activity of tPA: X-ray  
 RT crystal structure of single-chain human tPA.";  
 RL EMBO J. 16:4797-4805 (1997).  
 RN [19]  
 RP X-RAY CRYSTALLOGRAPHY (2.4 ÅNGSTRÖMS) OF KRINGLE 2.  
 RX MEDLINE=92118803; PubMed=1310033;  
 RA de Vos A., Ultsch M.H., Kelley R.F., Padmanabhan K., Tulinsky A.,  
 RA Westbrook M.L., Kossiakof A.A.,  
 RT "Crystal structure of the kringle 2 domain of tissue plasminogen  
 RT activator at 2.4-Å resolution.";  
 RL Biochemistry 31:270-279 (1992).  
 RN [20]  
 RP STRUCTURE BY NMR OF KRINGLE 2.  
 RX MEDLINE=90122799; PubMed=2589718;  
 RA Byeon I.-J.L., Kelley R.F., Llinas M.,  
 RT "1H NMR structural characterization of a recombinant kringle 2 domain  
 RT from human tissue-type plasminogen activator.";  
 RL Biochemistry 28:9350-9360 (1989).  
 RN [21]  
 RP STRUCTURE BY NMR OF KRINGLE 2.  
 RX MEDLINE=91200042; PubMed=1901789;  
 RA Byeon I.-J.L., Kelley R.F., Llinas M.,  
 RT "Kringle-2 domain of the tissue-type plasminogen activator. 1H-NMR  
 RT assignments and secondary structure.";  
 RL Eur. J. Biochem. 197:155-165 (1991).  
 RN [22]  
 RP STRUCTURE BY NMR OF KRINGLE 2.  
 RX MEDLINE=92106329; PubMed=1762144;  
 RA Byeon I.-J.L., Llinas M.,  
 RT "Solution structure of the tissue-type plasminogen activator kringle  
 RT 2 domain complexed to 6-aminohexanoic acid an antifibrinolytic  
 RT drug.";  
 RL J. Mol. Biol. 222:1035-1051 (1991).  
 RN [23]  
 Query Match 12.0%; Score 154.5; DB 1; Length 562;  
 Best Local Similarity 39.6%; Pred. No. 1.5e-05;  
 Matches 36; Conservative 6; Mismatches 36; Indels 13; Gaps 4;  
 QY 4 CFWDNGHLYREDQTSFAPGLRCINWLDQAQSLASAPVS-----GAGNHSYCRNPDE 54  
 DB 127 CYEDQGISYRGTWSTAESGAECTNW--NSSALAKPYSGRRPDAIRGLGNHNYCRNPD 184  
 QY 55 DPGRPWCYVSGEAGVPEKRCEDLRCPETTS 85  
 DB 185 DSK-PWCYVFP-KAGKYSSEFCSTFACSEGS 213  
 RESULT 3

TPA_RAT	TPA_RAT	STANDARD;	PRT;	559 AA.
ID	AC	PL9637;		
AC	DT	01-FEB-1991 (Rel. 17, Created)		
DT	DT	01-NOV-1991 (Rel. 20, Last sequence update)		
DT	DT	10-OCT-2003 (Rel. 42, Last annotation update)		
DE	DE	Tissue-type plasminogen activator precursor (EC 3.4.21.68) (tPA)		
DE	DE	(t-PA) (t-plasminogen activator).		
GN	PLAT.			
CS	Rattus norvegicus (Rat).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.			
ON	NCBI_TaxID=10116;			
OX	[1]			
RP	SEQUENCE FROM N.A.			
RP	MEDLINE=89170114; PubMed=3148445;			
RP	NY T.; Leonardson G.; Hsueh A.J.W.;			
RA	"Cloning and characterization of a cDNA for rat tissue-type			
RT	plasminogen activator.";			
RT	DNA 7:671-677(1988).			
RL	[2]			
RP	SEQUENCE FROM N.A.			
RP	MEDLINE=90130448; PubMed=2105315;			
RP	Peng P.; Ohlsson M., NY T.;			
RA	"The structure of the TATA-less rat tissue-type plasminogen activator			
RT	gene. Species-specific sequence divergences in the promoter predict			
RT	differences in regulation of gene expression.";			
RL	J. Biol. Chem. 265:2022-2027(1990).			
CC	-1- FUNCTION: Converts the abundant, but inactive, zymogen plasminogen			
CC	to plasmin by hydrolyzing a single Arg-Val bond in plasminogen. By			
CC	controlling plasmin-mediated proteolysis, it plays an important			
CC	role in tissue remodeling and degradation, in cell migration and			
CC	many other physiopathological events.			
CC	CATALYTIC ACTIVITY: Specific cleavage of Arg- -Val bond in			
CC	plasminogen to form plasmin.			
CC	-1- SUBUNIT: Heterodimer of chain A and chain B held by a disulfide			
CC	bond.			
CC	-1- SUBCELLULAR LOCATION: Secreted; extracellular.			
CC	-1- PM: THE SINGLE CHAIN, ALMOST FULLY ACTIVE ENZYME, CAN BE FURTHER			
CC	PROCESSED INTO A TWO-CHAIN FULLY ACTIVE FORM BY A CLEAVAGE AFTER			
CC	ARG-308 CATALYZED BY PLASMIN, TISSUE KALLIKREIN OR FACTOR XA.			
CC	-1- MISCELLANEOUS: Binds to the kringle structure of the fibrin A			
CC	chain. Binding to fibrin enhances its catalytic activity.			
CC	-1- SIMILARITY: Belongs to peptidase family S1.			
CC	-1- SIMILARITY: Contains 1 EGF-like domain.			
CC	-1- SIMILARITY: Contains 1 fibronectin type I domain.			
CC	-1- SIMILARITY: Contains 2 kringle domains.			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (See <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a>			
CC	or send an email to <a href="mailto:license@isb-sib.ch">license@isb-sib.ch</a> ).			
CC	-----			
DR	EMBL; M23697; AAA41812.1;			
DR	EMBL; M31197; AAA42261.1;			
DR	EMBL; M31185; AAA42261.1; JOINED.			
DR	EMBL; M31186; AAA42261.1; JOINED.			
DR	EMBL; M31187; AAA42261.1; JOINED.			
DR	EMBL; M31188; AAA42261.1; JOINED.			
DR	EMBL; M31189; AAA42261.1; JOINED.			
DR	EMBL; M31190; AAA42261.1; JOINED.			
DR	EMBL; M31191; AAA42261.1; JOINED.			
DR	EMBL; M31192; AAA42261.1; JOINED.			
DR	EMBL; M31193; AAA42261.1; JOINED.			
DR	EMBL; M31194; AAA42261.1; JOINED.			
DR	EMBL; M31195; AAA42261.1; JOINED.			
DR	EMBL; M31196; AAA42261.1; JOINED.			
DR	EMBL; A19618; CAA01482.1;			
DR	PIR; A35029; A35029.			
DR	HSP; P00750; 1RTF.			

Db 124 CFEQGITVGTWSTAENGAECINW--NSSALSQKPSARRPNAIKLGLNHNCRNPR 181

QY 55 DRPGPCWYSGEAGVPEKPCEDLRCE-----TTSQALPAFTTEIQEASEGPG 103

Db 182 DYK-PWCYVF-KAGKYTTFECSPACPKGFTEDCYGKGVYRGTHSFTT--SKASCLPW 237

QY 104 ADEQVQFAPANALPARSEA 122

Db 238 NSMILIGKTYTAWRANSOA 256

RESULT 4

HGFA\_MOUSE STANDARD; PRT; 653 AA.

AC Q9R098; Q9JKV4;

DT 16-OCT-2001 (Rel. 40, Created)

DT 16-OCT-2001 (Rel. 40, Last sequence update)

DT 15-MAR-2004 (Rel. 43, Last annotation update)

DE Hepatocyte growth factor activator precursor (EC 3.4.21.-) (HGF

DE activator) (HGFA).

GN HGFA.

OS Mus musculus (Mouse).

OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

OX NCBI\_TaxID=10090;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=BALB/c;

RA Itoh H., Kataoka H., Koono H.;

RT "Mouse hepatocyte growth factor activator.";

RL Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.

RN [2]

RP SEQUENCE FROM N.A.

RX MEDLINE=21226753; PubMed=1032833;

RA van Adelsberg J.S., Sehgal S., Kukes A., Brady C., Barasch J.,

RA Yang J., Huan J.;

RT "Activation of hepatocyte growth factor (HGF) by endogenous HGF

RT activator is required for metanephric kidney morphogenesis in

RT vitro.";

RL J. Biol. Chem. 276:15099-15106(2001).

CC -!- FUNCTION: Activates hepatocyte growth factor (HGF) by converting

CC it from a single chain to a heterodimeric form (By similarity).

CC -!- SUBUNIT: Dimer of a short chain and a long chain linked by a

CC disulfide bond (By similarity).

CC -!- SUBCELLULAR LOCATION: Secreted as an inactive single-chain

CC precursor and is then activated to a heterodimeric form (By

CC similarity).

CC -!- SIMILARITY: Belongs to peptidase family S1.

CC -!- SIMILARITY: Contains 2 EGF-like domains.

CC -!- SIMILARITY: Contains 1 fibronectin type I domain.

CC -!- SIMILARITY: Contains 1 fibronectin type II domain.

CC -!- SIMILARITY: Contains 1 kringle domain.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).

CC -----

CC EMBL; AF099017; AAF02489.1; -

CC EMBL; AF024724; AAF34712.1; -

CC HSPB; P00763; IDPO.

CC MEROPS; S01.228; -

CC MGD; MGI:1859281; Hgfac.

CC InterPro; IPR009003; Cys\_Ser\_trypsin.

CC InterPro; IPR000742; EGF\_2.

CC InterPro; IPR006209; EGF-like.

CC InterPro; IPR000083; Fibnactrl.

CC InterPro; IPR000562; FN\_Type\_II.

CC InterPro; IPR006210; IEGF.

CC InterPro; IPR000001; Kringle.

DR InterPro; IPR001254; Peptidase S1.

DR InterPro; IPR001314; Peptidase\_S1A.

DR Pfam; PF00008; EGF\_2.

DR Pfam; PF00039; fn1; 1.

DR Pfam; PF00040; fn2; 1.

DR Pfam; PF00051; kringle; 1.

DR PRINTS; PR00722; CHYMOTRYPSIN.

DR PRINTS; PR00013; FNTYPEII.

DR PRINTS; PR00018; KRINGLE.

DR ProDom; PD000995; FN\_Type\_II; 1.

DR ProDom; PD000395; Kringle; 1.

DR SMART; SM00181; EGF\_2.

DR SMART; SM00059; FN2; 1.

DR SMART; SM00130; KR; 1.

DR SMART; SM00020; Tryp\_Spc; 1.

DR PROSITE; PS00022; EGF\_1; 2.

DR PROSITE; PS01186; EGF\_2; 1.

DR PROSITE; PS50026; EGF\_3; 2.

DR PROSITE; PS01253; FIBRONECTIN\_1; 1.

DR PROSITE; PS00023; FIBRONECTIN\_2; 1.

DR PROSITE; PS00021; KRINGLE\_1; 1.

DR PROSITE; PS00020; KRINGLE\_2; 1.

DR PROSITE; PS50240; TRYPSIN\_DOM; 1.

DR PROSITE; PS00134; TRYPSIN\_HIS; 1.

DR PROSITE; PS00135; TRYPSIN\_SER; 1.

KW Hydrolase; Glycoprotein; Plasmin; Serine protease; Kringle; Signal;

KW EGF-like domain; Repeat; Zymogen.

FT SIGNAL 1 29 BY SIMILARITY.

FT PROPEP 30 369 CLEAVED IN ACTIVE FORM (BY SIMILARITY).

FT CHAIN 370 405 HEPATOCYTE GROWTH FACTOR ACTIVATOR SHORT

FT CHAIN 406 653 HEPATOCYTE GROWTH FACTOR ACTIVATOR LONG

FT CHAIN 105 145 FIBRONECTIN TYPE-II.

FT DOMAIN 157 195 EGF-LIKE 1.

FT DOMAIN 197 237 FIBRONECTIN TYPE-I.

FT DOMAIN 238 276 EGF-LIKE 2.

FT DOMAIN 283 364 KRINGLE.

FT DOMAIN 406 653 SERINE PROTEASE.

FT ACT\_SITE 445 445 CHARGE RELAY SYSTEM (BY SIMILARITY).

FT ACT\_SITE 495 495 CHARGE RELAY SYSTEM (BY SIMILARITY).

FT ACT\_SITE 596 596 CHARGE RELAY SYSTEM (BY SIMILARITY).

FT DISULFID 105 130 BY SIMILARITY.

FT DISULFID 119 145 BY SIMILARITY.

FT DISULFID 161 172 BY SIMILARITY.

FT DISULFID 166 183 BY SIMILARITY.

FT DISULFID 185 194 BY SIMILARITY.

FT DISULFID 199 227 BY SIMILARITY.

FT DISULFID 225 234 BY SIMILARITY.

FT DISULFID 242 253 BY SIMILARITY.

FT DISULFID 247 264 BY SIMILARITY.

FT DISULFID 266 275 BY SIMILARITY.

FT DISULFID 283 364 BY SIMILARITY.

FT DISULFID 304 346 BY SIMILARITY.

FT DISULFID 335 359 BY SIMILARITY.

FT DISULFID 392 519 INTERCHAIN (BY SIMILARITY).

FT DISULFID 430 446 BY SIMILARITY.

FT DISULFID 438 508 BY SIMILARITY.

FT DISULFID 533 602 BY SIMILARITY.

FT DISULFID 565 581 BY SIMILARITY.

FT DISULFID 592 620 BY SIMILARITY.

FT CARBOHYD 39 39 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 47 47 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 63 63 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 287 287 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 466 466 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 544 544 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CONFLICT 164 164 G -> W (IN REF. 2).

SC SEQUENCE 653 AA; 70567 MW; 88B4B20255DF7DC CRC64;

Query Match 11.6%; Score 150; DB 1; Length 653;

Best Local Similarity 32.6%; Pred. No. 4.2e-05;



[illegible]



```

FT DISULFID 202 230 BY SIMILARITY.
FT DISULFID 228 237 BY SIMILARITY.
FT DISULFID 245 256 BY SIMILARITY.
FT DISULFID 250 267 BY SIMILARITY.
FT DISULFID 269 278 BY SIMILARITY.
FT DISULFID 286 367 BY SIMILARITY.
FT DISULFID 307 349 BY SIMILARITY.
FT DISULFID 338 362 BY SIMILARITY.
FT DISULFID 394 521 INTERCHAIN (BY SIMILARITY).
FT DISULFID 432 448 BY SIMILARITY.
FT DISULFID 440 510 BY SIMILARITY.
FT DISULFID 535 604 BY SIMILARITY.
FT DISULFID 567 583 BY SIMILARITY.
FT DISULFID 594 622 BY SIMILARITY.
FT CARBOHYD 48 48 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 290 290 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 468 468 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 492 492 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 546 546 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CONFLICT 644 644 R -> O (IN REF. 2).
SQ SEQUENCE 655 AA; 70681 MW; 2CF72F1E1B862ED7 CRC64;

Query Match 11.4%; Score 146.5; DB 1; Length 655;
Best Local Similarity 36.9%; Pred. No. 8.3e-05;
Matches 41; Conservative 10; Mismatches 41; Indels 19; Gaps 6;

QY 4 CFWDNHLYREDQTSRPLGRCLNW-----LDAQSGLASAFVSGAGNHSYCRNPDED 55
Db 286 CFLGNGYGVGAVSTASGUSCLAWNSDLLYQELHVDTS-VGAALLGLGFHAYCRNPDD 344

QY 56 PRGFWCVSAGEAGVP-----EXRPECEDLCRPETTSQALPAFTTBIQF-ASEG 101
Db 345 ER-PWCYVVDXSALSWEYCRLEACSL----TRVQLSPDLLATLPEPAPSPG 390

RESULT 7
ID PLMN HUMAN STANDARD; PRT; 810 AA.
AC P00747;
DT 21-JUL-1986 (Rel. 01, Created)
DT 01-MAR-1989 (Rel. 10, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Plasminogen precursor (EC 3.4.21.7) [Contains: Angiostatin].
GN PLG.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A., AND VARIANT ASN-472.
RX MEDLINE=9020879; PubMed=2318848;
RA Petersen T.E., Martzen M.R., Ichinose A., Davie E.W.;
RT "Characterization of the gene for human plasminogen, a key proenzyme
RT in the fibrinolytic system."
RL J. Biol. Chem. 265:6104-6111(1990).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=87162490; PubMed=3030813;
RA Forsgren M., Raden B., Israelsson M., Larsson K., Heden L.-O.;
RT "Molecular cloning and characterization of a full-length cDNA clone
RT for human plasminogen."
RL FEBS Lett. 213:254-260(1987).
RN [3]
RP SEQUENCE FROM N.A., AND VARIANTS LYS-57; GLN-133; HIS-261; TRP-408;
RP ASN-472; VAL-494 AND TRP-523.
RA Rieder M.J., Armel T.Z., Carrington D.P., Ozuna M., Kuldanek S.A.,
RA Rajkumar N., Toth E.J., Yi Q., Nickerson D.A.;
RL Submitted (DEC-2002) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE OF 20-810, AND VARIANT ASN-472.
RA Sottrup-Jensen L., Petersen T.E., Magnusson S.;
RL Submitted (JUL-1977) to the PIR data bank.
RN [5]

```

```

RP RX SEQUENCE OF 292-810 FROM N.A.
RX MEDLINE=85023311; PubMed=6148961;
RA Malinowski D.P., Sadler J.E., Davie E.W.;
RT "Characterization of a complementary deoxyribonucleic acid coding for
RT human and bovine plasminogen."
RL Biochemistry 23:4243-4250(1984).
RN [6]
RP SEQUENCE OF 20-100.
RX MEDLINE=75093329; PubMed=122932;
RA Wiman B., Wallen P.;
RT "Structural relationship between 'glutamic acid' and 'lysine' forms
RT of human plasminogen and their interaction with the NH2-terminal
RT activation peptide as studied by affinity chromatography."
RN [7]
RP SEQUENCE OF 95-580; 581-626; 657-700 AND 732-810, AND VARIANT ASN-472.
RP Sottrup-Jensen L., Claeys H., Zajdel M., Petersen T.E., Magnusson S.;
RL (In) Davidson J.F., Rowan R.M., Samama M.M., Desnoyers P.C. (eds.);
RL Progress in chemical fibrinolysis and thrombolysis, pp.3:191-209,
RL Raven Press, New York (1978).
RN [8]
RP SEQUENCE OF 483-604.
RX MEDLINE=76043692; PubMed=126863;
RA Wiman B., Wallen P.;
RT "Amino-acid sequence of the cyanogen-bromide fragment from human
RT plasminogen that forms the linkage between the plasmin chains."
RL Eur. J. Biochem. 58:539-547(1975).
RN [9]
RP SEQUENCE OF 581-810.
RX MEDLINE=77225245; PubMed=142009;
RA Wiman B.;
RT "Primary structure of the B-chain of human plasmin."
RN [10]
RP ACTIVE SITE.
RX MEDLINE=73149248; PubMed=4694729;
RA Robbins K.C., Bernabe P., Arzadon L., Summaria L.;
RT "The primary structure of human plasminogen. II. The histidine loop
RT of human plasmin: light (B) chain active center histidine sequence."
RN [11]
RP ACTIVE SITE.
RX MEDLINE=69234739; PubMed=4240117;
RA Groskopf W.R., Summaria L., Robbins K.C.;
RT "Studies on the active center of human plasmin. Partial amino acid
RT sequence of a peptide containing the active center serine residue."
RN [12]
RP OMEGA-AMINOCARBOXYLIC ACID-BINDING SITES.
RX MEDLINE=82213905; PubMed=6919539;
RA Trexler M., Vali Z., Pathy L.;
RT "Structure of the omega-aminocarboxylic acid-binding sites of human
RT plasminogen. Arginine 70 and aspartic acid 56 are essential for
RT binding of ligand by kringle 4."
RL J. Biol. Chem. 257:7401-7406(1982).
RN [13]
RP FIBRIN AND OMEGA-AMINOCARBOXYLIC ACID BINDING SITES.
RX MEDLINE=85054794; PubMed=6094526;
RA Vali Z., Pathy L.;
RT "The fibrin-binding site of human plasminogen. Arginines 32 and 34
RT are essential for fibrin affinity of the kringle 1 domain."
RN [14]
RP PHOSPHORYLATION SITE SER-597.
RX MEDLINE=97345939; PubMed=9201958;
RA Wang H., Proxok M., Bretthauer R.K., Castellino F.J.;
RT "Serine-578 is a major phosphorylation locus in human plasma
RT plasminogen."
RL Biochemistry 36:8100-8106(1997).
RN [15]
RP CARBOHYDRATE-LINKAGE SITES.
RX MEDLINE=88185329; PubMed=3356193;
RA Marti T., Schaller J., Rickli E.E., Schmid K., Kamerling J.P.,

```

RA Gerwig G.J., van Halbeek H., Vliegthart J.F.;  
 RT "The N- and O-linked carbohydrate chains of human, bovine and porcine  
 RT plasminogen. Specificity in relation to sialylation and  
 RT fucosylation patterns.";  
 RL Eur. J. Biochem. 173:57-63(1988).  
 RN [16]  
 RP CARBOHYDRATE-LINKAGE SITE SER-268.  
 RX MEDLINE=97207306; PubMed=9054441;  
 RA Pirie-Shepherd S.R., Stevens R.D., Andon N.L., Enghild J.J.,  
 RA Pizzo S.V.;  
 RT "Evidence for a novel O-linked sialylated trisaccharide on Ser-248 of  
 RT human plasminogen 2.";  
 RL J. Biol. Chem. 272:7408-7411(1997).  
 RN [17]  
 RP CHARACTERIZATION OF ANGIOSTATIN, AND PARTIAL SEQUENCE.  
 RX MEDLINE=95042728; PubMed=7535077;  
 RA O'Reilly M.S., Holmgren L., Shing Y., Chen C., Rosenthal R.A.,  
 RA Moses M., Lane W.S., Cao E.H., Folkman J.;  
 RT "Angiostatin: a novel angiogenesis inhibitor that mediates the  
 RT suppression of metastases by a Lewis lung carcinoma.";  
 RL Cell 79:315-328(1994).  
 RN [18]  
 RP "A recombinant human angiostatin protein inhibits experimental primary  
 RT and metastatic cancer.";  
 RL Cancer Res. 57:1329-1334(1997).  
 RN [19]  
 RP X-RAY CRYSTALLOGRAPHY (1.9 ANGSTROMS) OF 374-461.  
 RX MEDLINE=92031502; PubMed=1657148;  
 RA Mulichak A.M., Tulinsky A., Ravichandran K.G.;  
 RT "Crystal and molecular structure of human plasminogen kringle 4  
 RT refined at 1.9-A resolution.";  
 RL Biochemistry 30:10576-10588(1991).  
 RN [20]  
 RP X-RAY CRYSTALLOGRAPHY (2.25 ANGSTROMS) OF 374-461.  
 RX MEDLINE=92031503; PubMed=1657149;  
 RA Wu T.-P., Padmanabhan K., Tulinsky A., Mulichak A.M.;  
 RT "The refined structure of the epsilon-aminocaproic acid complex of  
 RT human plasminogen kringle 4.";  
 RL Biochemistry 30:10589-10594(1991).  
 RN [21]  
 RP X-RAY CRYSTALLOGRAPHY (1.67 ANGSTROMS) OF 376-454.  
 RA Sec B., Yamano A., Whitlow M., Teeter M.M.;  
 RT "Structure of human plasminogen kringle 4 at 1.68 Angstrom and 277 K.  
 RT A possible structural role of disordered residues.";  
 RL Acta Crystallogr. D 53:169-178(1997).  
 RN [22]  
 RP X-RAY CRYSTALLOGRAPHY (2.1 ANGSTROMS) OF 102-181.  
 RX MEDLINE=96180681; PubMed=8611560;  
 RA Mathews I.I., Vanderhoff-Hanover P., Castellino F.J., Tulinsky A.;  
 RT "Crystal structures of the recombinant kringle 1 domain of human  
 RT plasminogen in complexes with the ligands epsilon-aminocaproic acid  
 RT and trans-4-(aminomethyl)cyclohexane-1-carboxylic Acid.";  
 RL Biochemistry 35:2567-2576(1996).  
 RN [23]  
 RP X-RAY CRYSTALLOGRAPHY (1.66 ANGSTROMS) OF 480-563.  
 RX MEDLINE=98198034; PubMed=9521645;  
 RA Chang Y., Mochalkin I., McCance S.G., Cheng B., Tulinsky A.,  
 RA Castellino F.J.;  
 RT "Structure and ligand binding determinants of the recombinant kringle  
 RT 5 domain of human plasminogen.";  
 RL Biochemistry 37:3258-3271(1998).  
 RN [24]  
 RP STRUCTURE BY NMR OF 96-184.  
 RX MEDLINE=94237157; PubMed=8181475;  
 RA Rejante M.R., Llinas M.;  
 RT "1H-NMR assignments and secondary structure of human plasminogen  
 RT kringle 1.";  
 RL Eur. J. Biochem. 221:927-937(1994).  
 RN [25]

RP STRUCTURE BY NMR OF 96-184.  
 RX MEDLINE=94237158; PubMed=8181476;  
 RA Rejante M.R., Llinas M.;  
 RT "Solution structure of the epsilon-aminohexanoic acid complex of  
 RT human plasminogen kringle 1.";  
 RL Eur. J. Biochem. 221:939-949(1994).  
 RN [26]  
 RP STRUCTURE BY NMR OF 183-354.  
 RX MEDLINE=96194156; PubMed=8652577;  
 RA Soehndel S., Hu C.-K., Marti D., Affolter M., Schaller J., Llinas M.,  
 RA Rickli E.B.;  
 RT "Recombinant gene expression and 1H NMR characteristics of the  
 RT kringle (2 + 3) supermodule: spectroscopic/functional individuality  
 RT of plasminogen kringle domains.";  
 RL Biochemistry 35:2357-2364(1996).  
 RN [27]  
 RP STRUCTURE BY NMR OF 374-461.  
 RX MEDLINE=90219023; PubMed=2157850;  
 RN [28]  
 RP Query Match 11.0%; Score 142; DB 1; Length 810;  
 Best Local Similarity 37.2%; Pred. No. 0.00025;  
 Matches 32; Conservative 8; Mismatches 34; Indels 12; Gaps 4;  
 QY 4 CFWNHGLHYREDQTSPPAGRLCLNWLDA-----QSGLASAPVSGAGNHSYCRNDEPRG 58  
 DB 103 CKTGKNGKRYGTMTSKNGITCKQWSSTSPHRPFRSPATHPSGL-ENYCRNPDPDQ 161  
 QY 59 PWCYVSGBAGVPEK---PCEDLRCE 82  
 DB 162 PWCYTD-----PEKGYDYCDILECEE 183  
 ID PLMN RAT STANDARD; PRT; 169 AA.  
 AC Q01177;  
 DT 01-APR-1993 (Rel. 25, Created)  
 DT 01-APR-1993 (Rel. 25, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Plasminogen (EC 3.4.21.7) (Fragment).  
 GN PLG.  
 OS Rattus norvegicus (Rat).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Rattus.  
 OX NCBI\_TaxID=10116;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Liver;  
 RX MEDLINE=91250378; PubMed=1645711;  
 RA Kanalas J.J., Makker S.P.;  
 RT "Identification of the rat Heymann nephritis autoantigen (GP330) as a  
 RT receptor site for plasminogen.";  
 RL J. Biol. Chem. 266:10825-10829(1991).  
 CC -!- FUNCTION: Plasmin dissolves the fibrin of blood clots and acts as  
 CC a proteolytic factor in a variety of other processes including  
 CC embryonic development, tissue remodeling, tumor invasion, and  
 CC inflammation; in ovulation it weakens the walls of the Graafian  
 CC follicle. It activates the urokinase-type plasminogen activator,  
 CC collagenases and several complement zymogens, such as C1 and C5.  
 CC It cleaves fibrin, fibronectin, thrombospondin, laminin and von  
 CC Willebrand factor.  
 CC -!- CATALYTIC ACTIVITY: Preferential cleavage: Lys-|-Xaa -> Arg-|-Xaa;  
 CC higher selectivity than trypsin. Converts fibrin into soluble  
 CC products.  
 CC -!- ENZYME REGULATION: Converted into plasmin by plasminogen  
 CC activators, both plasminogen and its activator being bound to  
 CC fibrin. Cannot be activated with streptokinase.  
 CC -!- MISCELLANEOUS: Plasmin is inactivated by alpha-2-antiplasmin  
 CC immediately after dissociation from the clot.  
 CC -!- SIMILARITY: Belongs to peptidase family S1. Plasminogen subfamily.  
 CC -!- SIMILARITY: Contains 5 kringle domains.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL; M62832; AAA41884.1; --  
CC PIR; A40522; A40522.  
CC HSP; P00747; IPMK.  
CC MEROPS; S01.233; --  
CC InterPro; IPR000001; Kringle.  
CC InterPro; IPR001254; Peptidase\_S1.  
CC InterPro; IPR003966; Peptidase\_S1A\_pr.  
CC Pfam; PF00051; kringle; 2.  
CC PRINTS; PR00018; KRINGLE.  
CC PRINTS; PR01505; PROTHROMBIN.  
CC ProDom; PD000395; Kringle; 2.  
CC SMART; SM00130; KR; 1.  
CC PROSITE; PS00021; KRINGLE\_1; 1.  
CC PROSITE; PS00070; KRINGLE\_2; 2.  
CC PROSITE; PS0240; TRYPSIN\_DOM; PARTIAL.  
CC PROSITE; PS00134; TRYPSIN\_HIS; PARTIAL.  
CC PROSITE; PS00135; TRYPSIN\_SER; PARTIAL.  
CC Hydroxylase; Serine protease; Plasma; Glycoprotein; Fibrinolysis;  
CC Tissue remodeling; Blood coagulation; Kringle; Repeat.  
CC NON\_TER 1  
CC DOMAIN 1  
CC FT DOMAIN <1 10 KRINGLE 3 (BY SIMILARITY).  
CC FT DOMAIN 34 112 KRINGLE 4 (BY SIMILARITY).  
CC FT DOMAIN 139 >169 KRINGLE 5 (BY SIMILARITY).  
CC FT DISULFID 34 112 BY SIMILARITY.  
CC FT DISULFID 55 95 BY SIMILARITY.  
CC FT DISULFID 83 107 BY SIMILARITY.  
CC NON\_TER 169  
CC SEQUENCE 169 AA; 18401 MW; 77A54214C49D010C CRC64;

Query Match 10.7%; Score 138; DB 1; Length 169;  
Best Local Similarity 31.8%; Pred. No. 9.6e-05;  
Matches 35; Conservative 13; Mismatches 48; Indels 14; Gaps 5;  
Qy 4 CFWDNGLHREDQTSAPAGRCRLNWLDA-----QSGLASAPVSGAGNHYCRNPDDPRG 58  
Db 34 CYQNGKSYRGTSTTTGKKOSWSWTPHSHKTPANFPDGL-EMNYCRNPDDQRG 92  
Qy 59 PWCYVSGEAGVPEK--PCEDLRCPETTSQALPAFTTIOEAGEGPGADE 106  
Db 93 PWCFTTD----PSVRYEYCNLRKCSGTGGV--AESAIVPQVSPAPTSE 136

RESULT 9  
ID UROK\_PAPCY STANDARD; PRT; 433 AA.  
AC P16227;  
DT 01-APR-1990 (Rel. 14, Created)  
DT 01-APR-1990 (Rel. 14, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE Urokinase-type plasminogen activator precursor (SC 3.4.21.73) (uPA)  
DE (U-plasminogen activator).  
GN PLAU.

OS Papio cynocephalus (Yellow baboon).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea;  
OC Cercopithecoidea; Papio.  
OX NCBI\_TaxID=9556;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Thoracic aorta;  
RX MEDLINE=90287734; PubMed=2113276;  
RA Au Y. P. T., Wang T. W., Clowes A. W.;  
RT "Nucleotide and deduced amino acid sequences of baboon urokinase-type  
RT plasminogen activator";  
RL Nucleic Acids Res. 18:3411-3411(1990).

CC -!- CATALYTIC ACTIVITY: Specific cleavage of Arg-|-Val bond in  
CC

CC plasminogen to form plasmin.  
CC -!- SUBUNIT: Found in high and low molecular mass forms. Each consists  
CC of two chains, A and B. The high molecular mass form contains a  
CC long chain A. Cleavage occurs after residue 155 in the low  
CC molecular mass form to yield a short A1 chain (By similarity).  
CC -!- SIMILARITY: Belongs to peptidase family S1.  
CC -!- SIMILARITY: Contains 1 EGF-like domain.  
CC -!- SIMILARITY: Contains 1 kringle domain.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL; X51935; CAA36200.1; --  
CC PIR; S14687; UKBAY.  
CC HSP; P00749; ILMW.  
CC MEROPS; S01.231; --  
CC InterPro; IPR009003; Cys\_Ser\_trypsin.  
CC InterPro; IPR006209; EGF\_Like.  
CC InterPro; IPR006210; IEGF.  
CC InterPro; IPR000001; Kringle.  
CC InterPro; IPR008293; Pept\_S1A\_uPA.  
CC InterPro; IPR001254; Peptidase\_S1.  
CC InterPro; IPR001314; Peptidase\_S1A.  
CC Pfam; PF00051; kringle; 1.  
CC Pfam; PF00089; trypsin; 1.  
CC PIRSF; PIRSF001144; Urk\_plasm act; 1.  
CC PRINTS; PR00722; CHYMOTRYPSIN.  
CC PRINTS; PR00018; KRINGLE.  
CC ProDom; PD000395; Kringle; 1.  
CC SMART; SM00181; EGF; 1.  
CC SMART; SM00130; KR; 1.  
CC SMART; SM00020; Tryp\_Spc; 1.  
CC PROSITE; PS00022; EGF\_1; 1.  
CC PROSITE; PS01186; EGF\_2; FALSE\_NEG.  
CC PROSITE; PS00026; EGF\_3; 1.  
CC PROSITE; PS00021; KRINGLE\_1; 1.  
CC PROSITE; PS00070; KRINGLE\_2; 1.  
CC PROSITE; PS0240; TRYPSIN\_DOM; 1.  
CC PROSITE; PS00134; TRYPSIN\_HIS; 1.  
CC PROSITE; PS00135; TRYPSIN\_SER; 1.  
CC Plasminogen activation; Hydrolase; Serine protease; Glycoprotein;  
CC Kringle; EGF-like domain; Zymogen; Signal.  
CC SIGNAL 1 20 POTENTIAL.  
CC CHAIN 21 433 UROKINASE-TYPE PLASMINOGEN ACTIVATOR.  
CC CHAIN 21 176 CHAIN A (BY SIMILARITY).  
CC CHAIN 155 176 SHORT A CHAIN (A1) (BY SIMILARITY).  
CC CHAIN 178 433 CHAIN B (BY SIMILARITY).  
CC DOMAIN 26 62 EGF-LIKE.  
CC DOMAIN 69 150 KRINGLE.  
CC DOMAIN 151 177 CONNECTING PEPTIDE.  
CC DOMAIN 178 433 SERINE PROTEASE.  
CC DISULFID 30 38 BY SIMILARITY.  
CC DISULFID 32 50 BY SIMILARITY.  
CC DISULFID 52 61 BY SIMILARITY.  
CC DISULFID 167 298 INTERCHAIN (BY SIMILARITY).  
CC DISULFID 208 224 BY SIMILARITY.  
CC DISULFID 216 287 BY SIMILARITY.  
CC DISULFID 315 384 BY SIMILARITY.  
CC DISULFID 347 363 BY SIMILARITY.  
CC DISULFID 374 402 BY SIMILARITY.  
CC ACT\_SITE 223 223 CHARGE RELAY SYSTEM.  
CC ACT\_SITE 274 274 CHARGE RELAY SYSTEM.  
CC ACT\_SITE 378 378 CHARGE RELAY SYSTEM.  
CC CARBOHYD 324 324 N-LINKED (GLCNAC... ) (BY SIMILARITY).  
CC SEQUENCE 433 AA; 48595 MW; 816D22DFDDC8792 CRC64;

Query Match 10.7%; Score 138; DB 1; Length 433;  
Best Local Similarity 32.7%; Pred. No. 0.00027;  
Qy 4 CFWDNGLHREDQTSAPAGRCRLNWLDA-----QSGLASAPVSGAGNHYCRNPDDPRG 58  
Db 34 CYQNGKSYRGTSTTTGKKOSWSWTPHSHKTPANFPDGL-EMNYCRNPDDQRG 92  
Qy 59 PWCYVSGEAGVPEK--PCEDLRCPETTSQALPAFTTIOEAGEGPGADE 106  
Db 93 PWCFTTD----PSVRYEYCNLRKCSGTGGV--AESAIVPQVSPAPTSE 136

RESULT 9  
ID UROK\_PAPCY STANDARD; PRT; 433 AA.  
AC P16227;  
DT 01-APR-1990 (Rel. 14, Created)  
DT 01-APR-1990 (Rel. 14, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE Urokinase-type plasminogen activator precursor (SC 3.4.21.73) (uPA)  
DE (U-plasminogen activator).  
GN PLAU.

OS Papio cynocephalus (Yellow baboon).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea;  
OC Cercopithecoidea; Papio.  
OX NCBI\_TaxID=9556;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Thoracic aorta;  
RX MEDLINE=90287734; PubMed=2113276;  
RA Au Y. P. T., Wang T. W., Clowes A. W.;  
RT "Nucleotide and deduced amino acid sequences of baboon urokinase-type  
RT plasminogen activator";  
RL Nucleic Acids Res. 18:3411-3411(1990).

CC -!- CATALYTIC ACTIVITY: Specific cleavage of Arg-|-Val bond in  
CC

```

DR PROSITE; PS50070; KRINGLE 2; 1.  

DR PROSITE; PS50240; TRYPSIN_DOM; 1.  

DR PROSITE; PS00134; TRYPSIN_HIS; 1.  

DR PROSITE; PS00135; TRYPSIN_SER; 1.  

KW Plasminogen activation; Hydrolase; Serine protease; Glycoprotein;  

FW Kringle; EGF-like domain; Zymogen; Signal.  

FT SIGNAL 1 20 BY SIMILARITY.  

FT CHAIN 21 442 UROKINASE-TYPE PLASMINOGEN ACTIVATOR.  

FT CHAIN A (BY SIMILARITY).  

FT CHAIN B (BY SIMILARITY).  

FT CHAIN 190 442  

FT CHAIN 65  

FT DOMAIN 29 65  

FT DOMAIN 72 153 KRWINGLE.  

FT DOMAIN 154 189 CONNECTING PEPTIDE.  

FT DOMAIN 190 442 SERINE PROTEASE.  

FT CARBOHYD 152 152 N-LINKED (GLCNAC. . ).  

FT DISULFID 33 41 BY SIMILARITY.  

FT DISULFID 35 53 BY SIMILARITY.  

FT DISULFID 55 64 BY SIMILARITY.  

FT DISULFID 179 310 INTERCHAIN (BY SIMILARITY).  

FT DISULFID 220 236 BY SIMILARITY.  

FT DISULFID 228 299 BY SIMILARITY.  

FT DISULFID 324 393 BY SIMILARITY.  

FT DISULFID 356 372 BY SIMILARITY.  

FT DISULFID 383 411 BY SIMILARITY.  

FT ACT_SITE 235 235 CHARGE RELAY SYSTEM.  

FT ACT_SITE 286 285 CHARGE RELAY SYSTEM.  

FT ACT_SITE 387 387 CHARGE RELAY SYSTEM.  

FT CONFLICT 241 281 Q -> H (IN REF. 1; CAA25806).  

FT CONFLICT 242 242 Q -> H (IN REF. 1; CAA26511).  

FT CONFLICT 288 288 A -> GS (IN REF. 1; CAA25806).  

SQ SEQUENCE 442 AA; 49116 MW; BE32CFE50132IEE CRC64;
Query Match 10.5%; Score 135.5; DB 1; Length 442;  

Best Local Similarity 36.9%; Pred. No. 0.00046;  

Matches 38; Conservative 7; Mismatches 37; Indels 21; Gaps 5;
QY 4 CFWDNGHLYREDOTSPAPGLRCNLWLDAGSL-----ASAPVS---GAGNHSYCRNPDEDP 56  

Db 72 CFEGNGSHYRGKANTNTGGPFCPLPWNSATVLLNTYHAHRFDALQLGLGKHNYCRNP-DNQ 130  

QY 57 RGPWCYVVS-----GEAGYP-----EKRPCEDLRCPETTSQ 86  

Db 131 RRPCYVQVGLKQIVQECWVPNCSCGGESHRRPAYDGKNPFSTPE 173
RESULT 11
URTB DESRO STANDARD; PRT; 431 AA.
ID URTE DESRO STANDARD; PRT; 431 AA.
AC P98121:  

DT 01-FEB-1996 (Rel. 33, Created)  

DT 01-FEB-1996 (Rel. 33, Last sequence update)  

DT 28-FEB-2003 (Rel. 41, Last annotation update)  

DE Salivary plasminogen activator beta precursor (EC 3.4.21.68) (DSPA  

beta).  

OS Desmodus rotundus (Vampire bat).  

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  

OC Mammalia; Eutheria; Chiroptera; Microchiroptera; Phyllostomidae;  

OC Desmodontinae; Desmodus.  

OX NCBI_TaxId=9430;  

RN [1]  

RP SEQUENCE FROM N.A.  

RX T7SSUE=Salivary gland;  

RX MEDLINE=92039036; PubMed=1937019;  

RA Kraetzschmar J., Haendler B., Langer G., Boidol W., Bringmann P.,  

RA Alagon A., Donner P., Schleuning W.D.;  

RT "The plasminogen activator family from the salivary gland of the  

vampire bat Desmodus rotundus: cloning and expression.";  

RL Gene 105:229-237(1991).  

RN [2]  

RP CHARACTERIZATION:  

RX MEDLINE=93393059; PubMed=1309059;  

RA Schleuning W.-D., Alagon A., Boidol W., Bringmann P., Petri T.,  

RA Kraetzschmar J., Haendler B., Langer G., Baldus B., Witt W.,

```

```

SQ SEQUENCE 431 AA; 48221 MW; 699B5E675B162CBF CRC64;

Query Match 10.4%; Score 134; DB 1; Length 431;
Best Local Similarity 38.8%; Pred. No. 0.00059;
Matches 27; Conservative 9; Mismatches 20; Indels 14; Gaps 3;

QY 4 CFWDNGHLYREDQTSAPGLRCLNLWDAOSGL-----ASAPVSGAGNHSYCRNPD 53
Db 82 CYKQDQGVYRGVTWTSESAQCINW--NSNLLTRRTYNGRRSDAITLGLGNHNYCRNPD 138
QY 54 EDPFGPKCYV 63
Db 139 NSNK-PWCYV 147

RESULT 12
URT2 DESRO
ID _URT2 DESRO STANDARD; PRT; 477 AA.
AC P15638;
DT 01-APR-1990 (Rel. 14, Created)
DT 01-APR-1996 (Rel. 33, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Salivary plasminogen activator alpha 2 precursor (EC 3.4.21.68) (DSPA
alpha-2) (BAT-PA) (T-plasminogen activator).
OS Desmodus rotundus (Vampire bat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Chiroptera; Microchiroptera; Phyllostomidae;
OC Desmodontinae; Desmodus.
OX NCBI_TaxID=9430;
RN [1]
SEQUENCE FROM N.A.
RP RP TISSUE=Salivary gland;
RX RX MEDLINE=92039036; PubMed=1937019;
RA Kraetzschmar J., Haendler B., Langer G., Boidol W., Bringmann P.,
Alagon A., Donner P., Schleuning W.D.;
RT "The plasminogen activator family from the salivary gland of the
vampire bat Desmodus rotundus: cloning and expression.";
RL Gene 105:229-237(1991).
RN [2]
SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RP RP TISSUE=Salivary gland;
RX RX MEDLINE=90036867; PubMed=2509450;
RA Gardell S.J., Duong L.T., Diehl R.E., York J.D., Hare T.R.,
Register R.B., Jacobs J.W., Dixon R.A.F., Friedman P.A.;
RT "Isolation, characterization, and cDNA cloning of a vampire bat
salivary plasminogen activator."
RL J. Biol. Chem. 264:17947-17952(1989).
RN [3]
CHARACTERIZATION.
RP RP MEDLINE=93393059; PubMed=1309059;
RA Schleuning W.-D., Alagon A., Boidol W., Bringmann P., Petri T.,
Kraetzschmar J., Haendler B., Langer G., Baldus B., Witt W.,
Donner P.;
RT "Plasminogen activators from the saliva of Desmodus rotundus (common
vampire bat): unique fibrin specificity.";
RL Ann. N.Y. Acad. Sci. 667:395-403(1992).
CC -!- FUNCTION: Probably essential to support the feeding habits of this
exclusively haematophagous animal. Probable potent thrombolytic
agent.
CC -!- CATALYTIC ACTIVITY: Specific cleavage of Arg-|-Val bond in
plasminogen to form plasmin.
CC -!- ENZYME REGULATION: Activity toward plasminogen is stimulated in
the presence of fibrin I.
CC -!- SUBUNIT: Monomer.
CC -!- DOMAIN: The fibronectin type-I domain mediates binding to fibrin,
and the kringle domain apparently mediates fibrin-induced
stimulation of activity.
CC -!- SIMILARITY: Belongs to peptidase family S1.
CC -!- SIMILARITY: Contains 1 EGF-like domain.
CC -!- SIMILARITY: Contains 1 fibronectin type I domain.
CC -!- SIMILARITY: Contains 1 kringle domain.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration

```

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC  
DR EMBL; M63988; AAA31593.1; -;  
DR EMBL; J05082; AAA31596.1; -;  
DR PIR; A34369; A34369;  
DR PIR; J05098; J05098;  
DR HSSP; P98119; IAS1;  
DR MEROPS; S01.232; -;  
DR InterPro; IPR009003; Cys\_Ser\_trypsin.  
DR InterPro; IPR006209; EGF\_Like.  
DR InterPro; IPR000083; Fibronectin.  
DR InterPro; IPR006210; IEGF.  
DR InterPro; IPR000001; Kringle.  
DR InterPro; IPR001254; Peptidase\_S1.  
DR InterPro; IPR001314; Peptidase\_S1A.  
DR Pfam; PF00008; EGF; 1.  
DR Pfam; PF00039; fn1; 1.  
DR Pfam; PF00051; kringle; 1.  
DR Pfam; PF00089; trypsin; 1.  
DR PRINTS; PR00722; CHYMOTRYPSIN.  
DR PRINTS; PR00018; KRINGLE.  
DR PRODOM; PD000395; Kringle; 1.  
DR SMART; SM00181; EGF; 1.  
DR SMART; SM00058; FN1; 1.  
DR SMART; SM00130; KR; 1.  
DR SMART; SM00020; TYP\_SPC; 1.  
DR PROSITE; PS00022; EGF\_1; 1.  
DR PROSITE; PS01186; EGF\_2; 1.  
DR PROSITE; PS00026; EGF\_3; 1.  
DR PROSITE; PS01253; FIBRONECTIN\_1; 1.  
DR PROSITE; PS00021; KRINGLE\_1; 1.  
DR PROSITE; PS00070; KRINGLE\_2; 1.  
DR PROSITE; PS00240; TRYPSIN\_DOM; 1.  
DR PROSITE; PS00134; TRYPSIN\_HIS; 1.  
DR PROSITE; PS00135; TRYPSIN\_SER; 1.  
KW Plasminogen activation; Hydrolase; Serine protease; Glycoprotein;  
KW Kringle; EGF-like domain; Signal; Multigene family.  
FT SIGNAL 1 36 POTENTIAL.  
FT CHAIN 37 477 SALIVARY PLASMINOGEN ACTIVATOR ALPHA 2.  
FT DOMAIN 40 82 FIBRONECTIN TYPE-I.  
FT DOMAIN 83 121 EGF-LIKE.  
FT DOMAIN 128 209 KRINGLE.  
FT ACT\_SITE 225 272 SERINE PROTEASE.  
FT ACT\_SITE 321 321 CHARGE RELAY SYSTEM (BY SIMILARITY).  
FT ACT\_SITE 428 428 CHARGE RELAY SYSTEM (BY SIMILARITY).  
FT DISULFID 42 72 BY SIMILARITY.  
FT DISULFID 70 79 BY SIMILARITY.  
FT DISULFID 87 98 BY SIMILARITY.  
FT DISULFID 92 109 BY SIMILARITY.  
FT DISULFID 111 120 BY SIMILARITY.  
FT DISULFID 128 209 BY SIMILARITY.  
FT DISULFID 149 191 BY SIMILARITY.  
FT DISULFID 180 204 BY SIMILARITY.  
FT DISULFID 214 345 BY SIMILARITY.  
FT DISULFID 257 273 BY SIMILARITY.  
FT DISULFID 265 334 BY SIMILARITY.  
FT DISULFID 359 434 BY SIMILARITY.  
FT DISULFID 391 407 BY SIMILARITY.  
FT DISULFID 424 452 BY SIMILARITY.  
FT CARBOHYD 185 185 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 398 398 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CONFLICT 403 403 N -> K (IN REF. 2).  
FT CONFLICT 417 417 Y -> H (IN REF. 2).  
FT CONFLICT 435 435 M -> R (IN REF. 2).  
SQ SEQUENCE 477 AA; 53719 MW; 17486555C0B5077C CRC64;

Query Match 10.4%; Score 134; DB 1; Length 477;

Best Local Similarity 38.6%; Pred. No. 0.00066;  
Matches 27; Conservative 9; Mismatches 20; Indels 14; Gaps 3;  
Qy 4 CFWDNGHLYREDTSPAPGLRCINWLDQAQSL-----ASAFVSGAGNHSYCRNPD 53  
Db 128 CYKDQGVYTGTTSTSESGACINW---NSNLLTRTYNGRRSDATLTGLGNHNYCRNPD 184  
Qy 54 EDPRGPMCYV 63  
Db 185 NNSK-PWCYV 193  
RESULT 13  
UROC HUMAN  
ID UROC HUMAN STANDARD; PRT; 431 AA.  
AC P00749; Q15844; Q15618; Q969W6;  
DT 21-JUL-1986 (Rel. 01, Created)  
DT 20-MAR-1987 (Rel. 04, Last sequence update)  
DT 10-OCT-2003 (Rel. 42, Last annotation update)  
DE Urokinase-type plasminogen activator precursor (EC 3.4.21.73) (uPA)  
DE (U-plasminogen activator).  
GN PLAU.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=85215647; PubMed=2987867;  
RA Riccio A., Grimaldi G., Verde P., Sebastio G., Boast S., Blasi F.;  
RT "The human urokinase-plasminogen activator gene and its promoter.";  
RL Nucleic Acids Res. 13:2759-2771(1985).  
RN [2]  
RP SEQUENCE FROM N.A.  
RA Holmes W.E., Pennica D., Blaber M., Rey M.W., Guenzler W.A.,  
RA Steffens G.J., Heyneker H.L.;  
RT "Cloning and expression of the gene for pro-urokinase in Escherichia  
RT coli.";  
RN [3]  
RP Biotechnology 3:923-929(1985).  
RP SEQUENCE FROM N.A.  
RX MEDLINE=86056954; PubMed=2415429;  
RA Nagai M., Hiramatsu R., Kaneda T., Hayasuke N., Arimura H.,  
RA Nishida M., Suyama T.;  
RT "Molecular cloning of cDNA coding for human preprourokinase.";  
RL Gene 36:183-188(1985).  
RN [4]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=85203359; PubMed=3898571;  
RA Jacobs P., Cravador A., Lorian R., Brockly F., Colau B., Chuchana P.,  
RA van Elsen A., Herzog A., Bollen A.;  
RT "Molecular cloning, sequencing, and expression in Escherichia coli of  
RT human preprourokinase cDNA.";  
RL DNA 4:139-146(1985).  
RN [5]  
RP SEQUENCE FROM N.A.  
RA Rieder M.J., Carrington D.P., Chung M.-W., Lee K.L., Poel C.L., Yi Q.,  
RA Nickerson D.A.;  
RT Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.  
RN [6]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Lung;  
RX MEDLINE=2238257; PubMed=12477932;  
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,  
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,  
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Heieh F.,  
RA Diachenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,  
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,  
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,  
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,  
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,



RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
RA Fahey J., Helton E., Kettman M., Madan A., Rodrigues S., Sanchez A.,  
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
RA Rodriguez A.C., Gaimwood J., Schmutz J., Myers R.M.,  
RA Butterfield Y.S.N., Krzywinaki M.I., Skalska U., Smailus D.E.,  
RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.,  
RT "Generation and initial analysis of more than 15,000 full-length  
RT human and mouse cDNA sequences."  
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).  
RN [7]  
RP SEQUENCE OF 66-431 FROM N.A.  
RX MEDLINE=84272706; PubMed=6589620;  
RA Verde P., Stoppelli M.P., Galeffi P., di Nocera P., Biasi F.,  
RT "Identification and primary sequence of an unspliced human urokinase  
RT poly(A)+ RNA."  
RL Proc. Natl. Acad. Sci. U.S.A. 81:4727-4731 (1984).  
RN [8]  
RP SEQUENCE OF 21-177.  
RX MEDLINE=83055084; PubMed=6754569;  
RA Gunzler W.A., Steffens G.J., Otting F., Kim S.-M.A., Frankus E.,  
RA Flohe L.,  
RT "The primary structure of high molecular mass urokinase from human  
RT urine. The complete amino acid sequence of the A chain."  
RL Hoppe-Seyler's Z. Physiol. Chem. 363:1155-1165 (1982).  
RN [9]  
RP SEQUENCE OF 156-176 AND 179-224.  
RX MEDLINE=93003608; PubMed=6749491;  
RA Schaller J., Nick H., Rickli E.B., Gillesen D., Lergier W.,  
RA Studer R.O.,  
RT "Human low-molecular-weight urinary urokinase. Partial  
RT characterization and preliminary sequence data of the two polypeptide  
RT chains."  
RL Eur. J. Biochem. 125:251-257 (1982).  
RN [10]  
RP SEQUENCE OF 158-410.  
RX MEDLINE=83055099; PubMed=6754572;  
RA Steffens G.J., Gunzler W.A., Otting F., Frankus E., Flohe L.,  
RT "The complete amino acid sequence of low molecular mass urokinase  
RT from human urine."  
RL Hoppe-Seyler's Z. Physiol. Chem. 363:1043-1058 (1982).  
RN [11]  
RP X-RAY CRYSTALLOGRAPHY (2.5 ANGSTROMS).  
RX MEDLINE=96000859; PubMed=8591045;  
RA Spraggon G., Phillips C., Nowak U.K., Ponting C.P., Saunders D.,  
RA Dobson C.M., Stuart D.I., Jones E.Y.,  
RT "The crystal structure of the catalytic domain of human  
RT urokinase-type plasminogen activator."  
RL Structure 3:681-691 (1995).  
RN [12]  
RP X-RAY CRYSTALLOGRAPHY (1.8 ANGSTROMS) OF 159-411.  
RX MEDLINE=20266327; PubMed=10805774;  
RA Sperl S., Jacob U., Arroyo de Prada N., Sturzebecher J., Wilhelm O.G.,  
RA Bode W., Magdolen V., Huber R., Moroder L.,  
RT "(4-aminomethyl)phenylguanidine derivatives as nonpeptidic highly  
RT selective inhibitors of human urokinase."  
RL Proc. Natl. Acad. Sci. U.S.A. 97:5113-5118 (2000).  
RN [13]  
RP STRUCTURE BY NMR.  
RX MEDLINE=89127526; PubMed=2536903;  
RA Oswald R.E., Bogusky M.J., Bamberger M., Smith R.A.G., Dobson C.M.,  
RT "Dynamics of the multidomain fibrinolytic protein urokinase from two-  
RT dimensional NMR."  
RL Nature 337:579-582 (1989).  
RN [14]  
RP STRUCTURE BY NMR OF 67-155.  
RX MEDLINE=93003110; PubMed=1327118;  
RA Li X., Smith R.A.G., Dobson C.M.,  
RT "Sequential 1H NMR assignments and secondary structure of the kringle  
RT domain from urokinase."  
RL Biochemistry 31:9562-9571 (1992).  
RN [15]  
RP STRUCTURE BY NMR OF 67-155.

RX MEDLINE=94149701; PubMed=8107091;  
RA Li X., Bokman A.M., Llinas M., Smith R.A.G., Dobson C.M.,  
RT "Solution structure of the kringle domain from urokinase-type  
RT plasminogen activator."  
RL J. Mol. Biol. 235:1548-1559 (1994).  
RN [16]  
RP VARIANT LEU-141.  
RX MEDLINE=96186279; PubMed=8652631;  
RA Yoshimoto M., Ushiyama Y., Sakai M., Tamaki S., Hara H., Takahashi K.,  
RA Sawasaki Y., Hanada K.,  
RT "Characterization of single chain urokinase-type plasminogen  
RT activator with a novel amino-acid substitution in the kringle  
RT structure."  
RL Biochim. Biophys. Acta 1293:83-89 (1996).  
RN [17]  
RP VARIANT LEU-141.  
RX MEDLINE=97218551; PubMed=9065988;  
RA Conne B., Berczy M., Belin D.,  
RT "Detection of polymorphisms in the human urokinase-type plasminogen  
RT activator gene."  
RL Thromb. Haemost. 77:434-435 (1997).  
RN [18]  
RP ERRATUM.  
RA Conne B., Berczy M., Belin D.,  
RL Thromb. Haemost. 78:973-973 (1997).  
RN [19]  
RP VARIANT LEU-141.  
RX MEDLINE=97337920; PubMed=914591;  
RA Turkmen B., Schmitt M., Schmalfeldt B., Trommler P., Hell W.,  
RA Creutzburg S., Graeff H., Magdolen V.,  
RT "Mutational analysis of the genes encoding urokinase-type plasminogen  
RT activator (uPA) and its inhibitor PAI-1 in advanced ovarian cancer."  
RL Electrophoresis 18:686-689 (1997).  
CC -!- FUNCTION: Potent plasminogen activator and is clinically used for  
CC therapy of thrombolytic disorders.  
CC -!- CATALYTIC ACTIVITY: Specific cleavage of Arg-|-Val bond in  
CC plasminogen to form plasmin.  
CC -!- SUBUNIT: Found in high and low molecular mass forms. Each consists  
CC of two chains, A and B. The high molecular mass form contains a  
CC long chain A. Cleavage occurs after residue 155 in the low  
CC molecular mass form to yield a short A1 chain.  
CC -!- PHARMACEUTICAL: Available under the name Abbokinase (Abbott). Used  
CC in Pulmonary Embolism (PE) to initiate fibrinolysis.  
CC -!- SIMILARITY: Belongs to peptidase family S1.  
CC -!- SIMILARITY: Contains 1 EGF-like domain.  
CC -!- SIMILARITY: Contains 1 kringle domain.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; X02419; CRA26268.1; -  
DR EMBL; M15476; AAA61253.1; -  
DR EMBL; D02444; BAA00175.1; -  
DR EMBL; D11143; BAA01919.1; -  
DR EMBL; X02760; CAR26535.1; -  
DR EMBL; AF377330; AAK53842.1; -  
DR EMBL; BC013575; AAH13575.1; -  
DR EMBL; K03226; AAC97138.1; -  
DR EMBL; K02286; AAA61252.1; -  
DR EMBL; A21571; CAA01559.1; -  
DR EMBL; A18397; CAA01390.1; -  
DR PIR; A00931; UKHU.  
DR PDB; 1KDU; 31-OCT-93.

Query Match 10.2%; Score 132; DB 1; Length 431;  
Best Local Similarity 32.7%; Pred. No. 0.00087;  
Matches 33; Conservative 13; Mismatches 33; Indels 22; Gaps 5;

QY 4 CFWNGHLYREDQTSAPGLRGLNWLDA-----QSGLASAPVSGAGNHSYCRNPDED 56  
Db 70 CYEGNGHLYRGKASTDTGRPCLPNSATVLOQTHAHSRDLQJLGLGKHNCRNPDP-NR 128  
QY 57 RGFVWCVYSGEAGVPEKRP-----CEDLRCPETTSQAL 88  
Db 129 RRFVWCVY--QVGL---KPLVQECVHWDCAHGKKPSSPPEL 164

RESULT 14  
HGFL\_MOUSE  
ID HGFL\_MOUSE STANDARD; PRT; 716 AA.  
AC P26928;  
DT 01-AUG-1992 (Rel. 23, Created)  
DT 01-AUG-1992 (Rel. 23, Last sequence update)  
DT 10-OCT-2003 (Rel. 42, Last annotation update)  
DE Hepatocyte growth factor-like protein precursor (Macrophage  
DE stimulatory protein) (MSP).  
GN MST1 OR HGFL.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=BALB/c; TISSUE=Liver;  
RX MEDLINE=92002017; PubMed=1832957;  
RA Friesner Degen S.J., Stuart L.A., Han S., Jamison C.S.;  
RT "Characterization of the mouse cDNA and gene coding for a hepatocyte  
RT growth factor-like protein: expression during development.";  
RL Biochemistry 30:9781-9791(1991).  
CC -!- FUNCTION: Probably has no proteolytic activity, since crucial AA  
CC characteristic of serine proteases catalytic sites are not  
CC conserved.  
CC -!- TISSUE SPECIFICITY: Liver. Lower levels in lung, placenta and  
CC adrenal.  
CC -!- DEVELOPMENTAL STAGE: Is expressed at low levels during gestation.  
CC Just before birth the level increases dramatically and remains  
CC stable afterwards.  
CC -!- PTM: MAY BE CLEAVED AFTER AA 488, TO YIELD A TWO-CHAIN MOLECULE  
CC HELD TOGETHER BY DISULFIDE BONDS, OR CLEAVED INTO TWO SEPARATE  
CC POLYPEPTIDES.  
CC -!- SIMILARITY: Belongs to peptidase family S1. Plasminogen subfamily.  
CC -!- SIMILARITY: Contains 4 kringle domains.

-----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; M74180; AAA50166.1; -.  
DR EMBL; M74181; AAA50167.1; -.  
DR PIR; A40332; A40332.  
DR HSP; P00747; 1KRN.  
DR MEROPS; S01.975; -.  
DR MGD; MGI:96080; Mstl.  
DR GO; GO:0007586; P:embryo implantation; IG.  
DR InterPro; IPR009003; Cys\_Ser\_trypsin.  
DR InterPro; IPR000001; Kringle.  
DR InterPro; IPR003014; PAN.  
DR InterPro; IPR003609; Pan app.  
DR InterPro; IPR001254; Peptidase S1.  
DR InterPro; IPR001314; Peptidase\_S1A.  
DR InterPro; IPR003966; Peptidase\_S1A\_pr.  
DR Pfam; PF00051; Kringle; 4.  
DR Pfam; PF00024; PAN; 1.  
DR Pfam; PF00089; trypsin; 1.  
DR PRINTS; PR00722; CHYMOTRYPSIN.  
DR PRINTS; PR00018; KRINGLE.  
DR PRINTS; PR01505; PROTHROMBIN.

ProDom: PD000395; Kringle; 4.  
SMART; SM00130; KR; 4  
SMART; SM00473; PAN\_AP; 1.  
DR SMART; SM00020; Tryp\_Spc; 1.  
DR PROSITE; PS00021; KRINGLE\_1; 4.  
DR PROSITE; PS00070; KRINGLE\_2; 4.  
DR PROSITE; PS00240; TRYPSIN\_DOM; 1.  
KW Kringle; Glycoprotein; Serine protease homolog; Repeat; Signal.  
FT SIGNAL 1 18 POTENTIAL.  
FT CHAIN 19 716 HEPATOCYTE GROWTH FACTOR-LIKE PROTEIN.  
FT DOMAIN 19 109 PAP.  
FT DOMAIN 110 186 KRINGLE 1.  
FT DOMAIN 191 268 KRINGLE 2.  
FT DOMAIN 292 370 KRINGLE 3.  
FT DOMAIN 379 457 KRINGLE 4.  
FT DOMAIN 489 716 SERINE PROTEASE-LIKE.  
FT DISULFID 56 78 BY SIMILARITY.  
FT DISULFID 60 66 BY SIMILARITY.  
FT DISULFID 110 186 BY SIMILARITY.  
FT DISULFID 131 169 BY SIMILARITY.  
FT DISULFID 157 181 BY SIMILARITY.  
FT DISULFID 191 268 BY SIMILARITY.  
FT DISULFID 194 333 INTERCHAIN (BY SIMILARITY).  
FT DISULFID 212 251 BY SIMILARITY.  
FT DISULFID 240 263 BY SIMILARITY.  
FT DISULFID 292 370 BY SIMILARITY.  
FT DISULFID 313 352 BY SIMILARITY.  
FT DISULFID 341 364 BY SIMILARITY.  
FT DISULFID 379 457 BY SIMILARITY.  
FT DISULFID 400 440 BY SIMILARITY.  
FT DISULFID 428 452 BY SIMILARITY.  
FT DISULFID 477 593 INTERCHAIN (BY SIMILARITY).  
FT DISULFID 512 528 BY SIMILARITY.  
FT DISULFID 607 672 BY SIMILARITY.  
FT DISULFID 637 651 BY SIMILARITY.  
FT DISULFID 662 690 BY SIMILARITY.  
FT CARBOHYD 72 72 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 173 173 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 305 305 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 620 620 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CONFLICT 19 19 P -> Q (IN GENOMIC SEQUENCE).  
SQ SEQUENCE 716 AA; 80588 MW; BECE02EF85213ACC CRC64;

Query Match 10.1%; Score 130.5; DB 1; Length 716;  
Best Local Similarity 30.7%; Pred. No. 0.0021;  
Matches 35; Conservative 7; Mismatches 37; Indels 35; Gaps 5;

QY 4 CFWNGHLYREDQTSAPGLRCLNW----LDAQSLASAPVSGAGNHSYCRNPDEDPRGPW 60  
Db 110 CIMDNGSVYRGTVARTAGGLPCQAWRRFPNDHKYTPKNGL-BEENFCRNPDPGRGPW 168  
QY 61 CYVS-----GEAGVPEK-RPCE--DLRCPET 83  
Db 169 CYTNRSVRFQSGIKYTCREAVCVLNGEDYRGVDVTEGRCQRWDLQHPHS 222

RESULT 15  
ID PLMN\_ERIEU STANDARD; PRT; 810 AA.  
AC Q29485;  
DT 01-NOV-1997 (Rel. 35, Created)  
DT 01-NOV-1997 (Rel. 35, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE Plasminogen precursor (EC 3.4.21.7).  
GN PLG.  
OS Erinaceus europaeus (Western European hedgehog).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Insectivora; Erinaceidae; Erinaceinae; Erinaceus.  
OX NCBI\_TaxID=9365;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Liver;  
RC MEDLINE=96025778; PubMed=7592597;  
RX

RA Lawn R.M., Boonmark N.W., Schwartz K., Lindahl G.E., Wade D.P.,  
RA Byrne C.D., Fong K.J., Meer K., Paddy L.;  
RT "The recurring evolution of lipoprotein(a)".  
RT hedgehog apolipoprotein(a)".  
RL J. Biol. Chem. 270:24004-24009 (1995).  
RN [2]  
RP REVISIONS.  
RA Lawn R.M.;  
RL Submitted (FEB-1997) to the EMBL/GenBank/DBJ databases.  
CC -!- FUNCTION: Plasmin dissolves the fibrin of blood clots and acts as  
CC a proteolytic factor in a variety of other processes including  
CC embryonic development, tissue remodeling, tumor invasion, and  
CC inflammation; in ovulation it weakens the walls of the Graafian  
CC follicle. It activates the urokinase-type plasminogen activator,  
CC collagenases and several complement zymogens, such as C1 and C5.  
CC It cleaves fibrin, fibronectin, thrombospondin, laminin and von  
CC Willebrand factor.  
CC -!- CATALYTIC ACTIVITY: Preferential cleavage: Lys-|-Xaa > Arg-|-Xaa;  
CC higher selectivity than trypsin. Converts fibrin into soluble  
CC products.  
CC -!- ENZYME REGULATION: Converted into plasmin by plasminogen  
CC activators, both plasminogen and its activator being bound to  
CC fibrin. Cannot be activated with streptokinase  
CC -!- MISCELLANEOUS: Plasmin is inactivated by alpha-2-antiplasmin  
CC immediately after dissociation from the clot.  
CC -!- SIMILARITY: Belongs to peptidase family S1. Plasminogen subfamily.  
CC -!- SIMILARITY: Contains 5 kringle domains.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
CC -----  
CC EMBL; U33171; RAC48717.1; -;  
CC PIR; I46260; I46260.  
CC HSSP; P00747; LPMK.  
CC MEROPS; S01.233; -;  
CC InterPro; IPR009003; Cys Ser trypsin.  
CC InterPro; IPR000001; Kringle.  
CC InterPro; IPR003014; PAN.  
CC InterPro; IPR003609; Pan app.  
CC InterPro; IPR001254; Peptidase S1.  
CC InterPro; IPR001314; Peptidase\_S1A.  
CC InterPro; IPR003966; Peptidase\_S1A\_pr.  
CC Pfam; PF00051; Kringle; 5.  
CC Pfam; PF00024; PAN; 1.  
CC Pfam; PF00089; trypsin; 1.  
CC PRINTS; PR00722; CHYMOTRYPSIN.  
CC PRINTS; PR00018; KRINGLE.  
CC PRINTS; PR01505; PROTHROMBIN.  
CC PRODOM; PD000395; Kringle; 5.  
CC SMART; SM00130; KR; 5.  
CC SMART; SM00473; PAN AP; 1.  
CC SMART; SM00020; TRYD SPc; 1.  
CC PROSITE; PS00021; KRINGLE 1; 5.  
CC PROSITE; PS00070; KRINGLE 2; 5.  
CC PROSITE; PS00240; TRYPSIN\_DOM; 1.  
CC PROSITE; PS00134; TRYPSIN\_HIS; 1.  
CC PROSITE; PS00135; TRYPSIN\_SER; 1.  
CC Hydrolase; Serine protease; Plasma; Glycoprotein; Fibrinolysis;  
CC Tissue remodeling; Blood coagulation; Kringle; Zymogen; Repeat;  
CC Signal.  
CC SIGNAL  
CC 1 19 BY SIMILARITY.  
CC CHAIN 20 910 PLASMINOGEN.  
CC CHAIN 20 582 PLASMIN HEAVY CHAIN A (BY SIMILARITY).  
CC CHAIN 593 910 PLASMIN LIGHT CHAIN B (BY SIMILARITY).  
CC DOMAIN 583 910 SERINE PROTEASE.  
CC DOMAIN 103 181 KRINGLE 1.  
CC DOMAIN 185 262 KRINGLE 2.  
CC DOMAIN 275 352 KRINGLE 3.

FT DOMAIN 379 456 KRINGLE 4.  
FT DOMAIN 482 561 KRINGLE 5.  
FT ACT\_SITE 622 622 CHARGE RELAY SYSTEM.  
FT ACT\_SITE 665 665 CHARGE RELAY SYSTEM.  
FT ACT\_SITE 760 760 CHARGE RELAY SYSTEM.  
FT CARBOHYD 339 339 N-LINKED (GLCNAC...) (POTENTIAL).  
SQ SEQUENCE 810 AA; 90902 MW; 8E75780946017A16 CRC64;  
Query Match 10.0%; Score 129; DB 1; Length 810;  
Best Local Similarity 34.1%; Pred No 0.0032;  
Matches 31; Conservative 7; Mismatches 31; Indels 22; Gaps 4;  
QY 4 CFWDNGHLYREDQTSAPAGRLCLNWLDAQGLASAP-----VSGAGNHSYCRNPD 53  
Db 103 CKVNGKYYRGTVSKTKTGLTCQKW-----SAETPHKPRFSPDENPSEGLDQNYCRNPD 156  
QY 54 EDPRGPKCYVSGEAGVPEKR--PCEDLRQPE 82  
Db 157 NDPKGPWCYTMD-----PEVRYEYCEIIQCED 183  
Search completed: March 17, 2004, 07:04:25  
Job time : 10.7712 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 17, 2004, 06:59:03 ; Search time 44.2961 Seconds  
(without alignments)  
1723.750 Million cell updates/sec

Title: US-10-057-951-2\_COPY\_22\_263

Perfect score: 1289  
Sequence: 1 SGCFWDNCHLYREDQTSFA.....PVDPQEGSTPLMQAGTQCA 242

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

SPTREMBL\_25:\*  
1: sp\_archaea:\*  
2: sp\_bacteria:\*  
3: sp\_fungi:\*  
4: sp\_human:\*  
5: sp\_invertebrate:\*  
6: sp\_mammal:\*  
7: sp\_mhc:\*  
8: sp\_organelle:\*  
9: sp\_phase:\*  
10: sp\_plant:\*  
11: sp\_rodent:\*  
12: sp\_virus:\*  
13: sp\_vertebrate:\*  
14: sp\_unclassified:\*  
15: sp\_rvirus:\*  
16: sp\_bacteriap:\*  
17: sp\_archaeap:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1289	100.0	263	4 Q96FE7	Q96fe7 homo sapien
2	1286	99.8	263	4 O00318	O00318 homo sapien
3	1275	98.9	263	4 Q8NCJ9	Q8ncj9 homo sapien
4	1028.5	79.8	264	11 Q7TMJ8	Q7tmj8 mus musculus
5	1024.5	79.5	264	11 Q811Z2	Q811z2 mus musculus
6	935	72.5	234	4 Q86VW2	Q86vw2 homo sapien
7	734.5	57.0	213	11 Q811Z3	Q811z3 mus musculus
8	611	47.4	263	13 Q7SXB3	Q7sxb3 brachydanio
9	167	13.0	562	6 Q8SQ23	Q8sq23 sus scrofa
10	154.5	12.0	291	4 Q7Z7N2	Q7z7n2 homo sapien
11	154.5	12.0	516	4 Q9BV99	Q9bv99 homo sapien
12	154.5	12.0	562	4 Q86VK8	Q86vk8 homo sapien
13	150	11.5	653	11 Q8VCS4	Q8vcs4 mus musculus
14	145.5	11.3	564	6 Q8MKB1	Q8mb1 oryctolagus
15	142	11.0	810	4 Q15146	Q15146 homo sapien
16	137	10.6	391	5 Q86PQ9	Q86pq9 cryptospori

17	136.5	10.6	420	13 Q90504	Q90504 eptatretus
18	136	10.6	421	13 Q8AXX3	Q8axx3 xenopus lae
19	135	10.5	385	5 Q25101	Q25101 herdmania m
20	135	10.5	812	11 Q9ROW3	Q9row3 rattus norv
21	134.5	10.4	704	13 Q90865	Q90865 gallus gall
22	133.5	10.4	716	11 P70521	P70521 rattus norv
23	131.5	10.2	313	13 Q9PU78	Q9pu78 crocodylus
24	131	10.2	154	4 Q96SE8	Q96se8 homo sapien
25	131	10.2	608	13 Q9PTW7	Q9ptw7 struthio ca
26	130.5	10.1	709	13 Q7ZTN9	Q7ztn9 xenopus lae
27	130.5	10.1	716	11 Q91XG8	Q91xg8 mus musculu
28	130.5	10.1	717	13 P70006	P70006 xenopus lae
29	129	10.0	90	4 Q8NG20	Q8ng20 homo sapien
30	129	10.0	616	6 Q97507	Q97507 sus scrofa
31	128.5	10.0	806	6 Q18783	Q18783 macropus eu
32	127	9.9	157	6 Q9TVAS	Q9tv8 bos taurus
33	127	9.9	395	4 Q9BZW1	Q9bw1 homo sapien
34	126.5	9.8	433	6 Q8MIL0	Q8mil0 oryctolagus
35	126.5	9.8	433	6 Q8MHV7	Q8mhv7 oryctolagus
36	126.5	9.8	728	11 Q8CSG5	Q8cs5 mus musculus
37	126	9.8	716	13 Q91691	Q91691 xenopus lae
38	125	9.7	622	4 Q7Z7F3	Q7z7f3 homo sapien
39	125	9.7	730	6 Q867B7	Q867b7 canis fami
40	124	9.6	429	13 Q8AVB0	Q8avb0 brachydanio
41	123	9.5	728	6 Q9BH09	Q9bh09 felis silve
42	120	9.3	612	13 Q804W7	Q804w7 fugu rubrip
43	119.5	9.3	560	4 Q14520	Q14520 homo sapien
44	119	9.2	231	11 Q8C6L2	Q8c6l2 mus musculu
45	119	9.2	334	6 Q46507	Q46507 papio hamad

#### ALIGNMENTS

RESULT 1

Q96FE7 PRELIMINARY; PRT; 263 AA.  
ID Q96FE7  
AC Q96FE7;  
DT 01-DEC-2001 (TREMBlrel. 19, Created)  
DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)  
DT 01-JUN-2003 (TREMBlrel. 24, Last annotation update)  
DE Hypothetical protein (HGFL(L) protein).  
GN HGFL.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Brain;  
RA Strausberg R.;  
RL Submitted (JUL-2001) to the EMBL/GenBank/DBSJ databases.  
RN [2]  
RP SEQUENCE FROM N.A.  
RA Chiang H., Chang M.;  
RL Submitted (JUL-2002) to the EMBL/GenBank/DBSJ databases.  
CC !- SIMILARITY: CONTAINS 1 KRINGLE DOMAIN.  
DR EMBL; BC011049; AAH11049.1; -.  
DR EMBL; AF528080; AAC33763.1; -.  
DR InterPro; IPR000001; Kringle.  
DR Pfam; PF00051; kringle; 1.  
DR PRINTS; P300018; KRINGLE.  
DR ProDom; PD000395; Kringle; 1.  
DR SMART; SM00130; KR; 1.  
DR PROSITE; PS00021; KRINGLE\_1; 1.  
DR PROSITE; PS00070; KRINGLE\_2; 1.  
KW Hypothetical protein; Glycoprotein; Kringle.  
SQ SEQUENCE 263 AA; 28234 MW; 197C3EE888FA242 CRC64;

Query Match 100.0%; Score 1289; DB 4; Length 263;

Best Local Similarity 100.0%; Pred. No. 4.5e-111; Indels 0; Gaps 0;  
Matches 242; Conservative 0; Mismatches 0;

QY 1 SGGCFWNGHLYREDQTSAPGLRCLNWLDAQSGLASAPVSGAGNHSYCRNPDEDPGFW 60  
DB 22 SGGCFWNGHLYREDQTSAPGLRCLNWLDAQSGLASAPVSGAGNHSYCRNPDEDPGFW 81  
QY 61 CYVSGEAGVPEKPCEDLRCPETTSQALPAFTTIOEASEGPGADEVQVFAPANALPARS 120  
DB 82 CYVSGEAGVPEKPCEDLRCPETTSQALPAFTTIOEASEGPGADEVQVFAPANALPARS 141  
QY 121 EAAAVQPVIGISQVRMNSKEKKDLGTLGYVLGITMMVIIIAIGAGIILGYSYKRGKDLK 180  
DB 142 EAAAVQPVIGISQVRMNSKEKKDLGTLGYVLGITMMVIIIAIGAGIILGYSYKRGKDLK 201  
QY 181 EQHDKQVCEREMORITPLSAFTNPTCEIVDEKTVVHTSQTVPDQEGSTPLMGQAGTP 240  
DB 202 EQHDKQVCEREMORITPLSAFTNPTCEIVDEKTVVHTSQTVPDQEGSTPLMGQAGTP 261  
QY 241 GA 242  
DB 262 GA 263  
RESULT 2  
O00318 PRELIMINARY; PRT; 263 AA.  
ID O00318  
AC O00318  
DT 01-JUL-1997 (TrEMBLrel. 04, Created)  
DT 01-JUL-1997 (TrEMBLrel. 04, Last sequence update)  
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
DE WUGSC:DJ51SN1.2 protein.  
GN WUGSC:DJ51SN1.2  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Du Z., Scheet P., Harper M.;  
RT "The sequence of H. sapiens PAC clone RP3-51SN1.";  
RL Submitted (MAY-1997) to the EMBL/GenBank/DBJ databases.  
[2]  
RP SEQUENCE FROM N.A.  
RA Waterston R.;  
RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.  
CC -!- SIMILARITY: CONTAINS 1 KRINGLE DOMAIN.  
DR EMBL; AK074688; BAC1140.1; -.  
DR HSSE; P00749; 1KDU.  
DR InterPro; IPR000001; Kringle.  
DR Pfam; PF00051; Kringle; 1.  
DR PRINTS; PR00018; KRINGLE.  
DR ProDom; PD000395; Kringle; 1.  
DR SMART; SM00130; KR; 1.  
DR PROSITE; PS00021; KRINGLE\_1; 1.  
DR PROSITE; PS00070; KRINGLE\_2; 1.  
KW Glycoprotein; Kringle.  
SQ SEQUENCE 263 AA; 28248 MW; 197C3EE8E54A242 CRC64;  
Query Match 99.8%; Score 1286; DB 4; Length 263;  
Best Local Similarity 99.6%; Pred. No. 8.6e-111; Indels 0; Gaps 0;  
Matches 241; Conservative 1; Mismatches 0;  
QY 1 SGGCFWNGHLYREDQTSAPGLRCLNWLDAQSGLASAPVSGAGNHSYCRNPDEDPGFW 60  
DB 22 SGGCFWNGHLYREDQTSAPGLRCLNWLDAQSGLASAPVSGAGNHSYCRNPDEDPGFW 81  
QY 61 CYVSGEAGVPEKPCEDLRCPETTSQALPAFTTIOEASEGPGADEVQVFAPANALPARS 120  
DB 82 CYVSGEAGVPEKPCEDLRCPETTSQALPAFTTIOEASEGPGADEVQVFAPANALPARS 141  
QY 121 EAAAVQPVIGISQVRMNSKEKKDLGTLGYVLGITMMVIIIAIGAGIILGYSYKRGKDLK 180  
DB 142 EAAAVQPVIGISQVRMNSKEKKDLGTLGYVLGITMMVIIIAIGAGIILGYSYKRGKDLK 201  
QY 181 EQHDKQVCEREMORITPLSAFTNPTCEIVDEKTVVHTSQTVPDQEGSTPLMGQAGTP 240  
DB 202 EQHDKQVCEREMORITPLSAFTNPTCEIVDEKTVVHTSQTVPDQEGSTPLMGQAGTP 261  
QY 241 GA 242  
DB 262 GA 263  
RESULT 3  
O00318 PRELIMINARY; PRT; 263 AA.  
ID O00318  
AC O00318  
DT 01-OCT-2002 (TrEMBLrel. 22, Created)  
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)  
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)  
DE Hypothetical protein FLJ90207.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Mammary gland;  
RA Isogai T., Ota T., Nishikawa T., Hayashi K., Otsuki T., Sugiyama T.,  
RA Suzuki Y., Nagai K., Sugano S., Ishii S., Kawai-Hio Y., Saito K.,  
RA Yamamoto J., Wakamatsu A., Nakamura Y., Kojima S., Nagahari K.,  
RA Masuho Y., Ono T., Okano K., Yoshikawa Y., Aotsuka S., Sasaki N.,  
RA Hattori A., Okumura K., Iwayanagi T., Ninomiya K.;  
RT "NEDO human cDNA sequencing project";  
RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.  
CC -!- SIMILARITY: CONTAINS 1 KRINGLE DOMAIN.  
DR EMBL; AK074688; BAC1140.1; -.  
DR InterPro; IPR000001; Kringle.  
DR Pfam; PF00051; Kringle; 1.  
DR PRINTS; PR00018; KRINGLE.  
DR ProDom; PD000395; Kringle; 1.  
DR SMART; SM00130; KR; 1.  
DR PROSITE; PS00021; KRINGLE\_1; 1.  
DR PROSITE; PS00070; KRINGLE\_2; 1.  
KW Hypothetical protein; Glycoprotein; Kringle.  
SQ SEQUENCE 263 AA; 28104 MW; 73A9294D5142C6C3 CRC64;  
Query Match 98.9%; Score 1275; DB 4; Length 263;  
Best Local Similarity 99.2%; Pred. No. 9e-110;  
Matches 240; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
QY 1 SGGCFWNGHLYREDQTSAPGLRCLNWLDAQSGLASAPVSGAGNHSYCRNPDEDPGFW 60  
DB 22 SGGCFWNGHLYREDQTSAPGLRCLNWLDAQSGLASAPVSGAGNHSYCRNPDEDPGFW 81  
QY 61 CYVSGEAGVPEKPCEDLRCPETTSQALPAFTTIOEASEGPGADEVQVFAPANALPARS 120  
DB 82 CYVSGEAGVPEKPCEDLRCPETTSQALPAFTTIOEASEGPGADEVQVFAPANALPARS 141  
QY 121 EAAAVQPVIGISQVRMNSKEKKDLGTLGYVLGITMMVIIIAIGAGIILGYSYKRGKDLK 180  
DB 142 EAAAVQPVIGISQVRMNSKEKKDLGTLGYVLGITMMVIIIAIGAGIILGYSYKRGKDLK 201  
QY 181 EQHDKQVCEREMORITPLSAFTNPTCEIVDEKTVVHTSQTVPDQEGSTPLMGQAGTP 240  
DB 202 EQHDKQVCEREMORITPLSAFTNPTCEIVDEKTVVHTSQTVPDQEGSTPLMGQAGTP 261  
QY 241 GA 242  
DB 262 GA 263  
RESULT 4  
O00318 PRELIMINARY; PRT; 264 AA.  
ID O00318  
AC O00318  
DT 01-OCT-2003 (TrEMBLrel. 25, Created)  
DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)

DB 202 EQHDKQVCEREMORITPLSAFTNPTCEIVDEKTVVHTSQTVPDQEGSTPLMGQAGTP 261  
QY 241 GA 242  
DB 262 GA 263  
RESULT 3  
O00318 PRELIMINARY; PRT; 263 AA.  
ID O00318  
AC O00318  
DT 01-OCT-2002 (TrEMBLrel. 22, Created)  
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)  
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)  
DE Hypothetical protein FLJ90207.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Mammary gland;  
RA Isogai T., Ota T., Nishikawa T., Hayashi K., Otsuki T., Sugiyama T.,  
RA Suzuki Y., Nagai K., Sugano S., Ishii S., Kawai-Hio Y., Saito K.,  
RA Yamamoto J., Wakamatsu A., Nakamura Y., Kojima S., Nagahari K.,  
RA Masuho Y., Ono T., Okano K., Yoshikawa Y., Aotsuka S., Sasaki N.,  
RA Hattori A., Okumura K., Iwayanagi T., Ninomiya K.;  
RT "NEDO human cDNA sequencing project";  
RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.  
CC -!- SIMILARITY: CONTAINS 1 KRINGLE DOMAIN.  
DR EMBL; AK074688; BAC1140.1; -.  
DR InterPro; IPR000001; Kringle.  
DR Pfam; PF00051; Kringle; 1.  
DR PRINTS; PR00018; KRINGLE.  
DR ProDom; PD000395; Kringle; 1.  
DR SMART; SM00130; KR; 1.  
DR PROSITE; PS00021; KRINGLE\_1; 1.  
DR PROSITE; PS00070; KRINGLE\_2; 1.  
KW Hypothetical protein; Glycoprotein; Kringle.  
SQ SEQUENCE 263 AA; 28104 MW; 73A9294D5142C6C3 CRC64;  
Query Match 98.9%; Score 1275; DB 4; Length 263;  
Best Local Similarity 99.2%; Pred. No. 9e-110;  
Matches 240; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
QY 1 SGGCFWNGHLYREDQTSAPGLRCLNWLDAQSGLASAPVSGAGNHSYCRNPDEDPGFW 60  
DB 22 SGGCFWNGHLYREDQTSAPGLRCLNWLDAQSGLASAPVSGAGNHSYCRNPDEDPGFW 81  
QY 61 CYVSGEAGVPEKPCEDLRCPETTSQALPAFTTIOEASEGPGADEVQVFAPANALPARS 120  
DB 82 CYVSGEAGVPEKPCEDLRCPETTSQALPAFTTIOEASEGPGADEVQVFAPANALPARS 141  
QY 121 EAAAVQPVIGISQVRMNSKEKKDLGTLGYVLGITMMVIIIAIGAGIILGYSYKRGKDLK 180  
DB 142 EAAAVQPVIGISQVRMNSKEKKDLGTLGYVLGITMMVIIIAIGAGIILGYSYKRGKDLK 201  
QY 181 EQHDKQVCEREMORITPLSAFTNPTCEIVDEKTVVHTSQTVPDQEGSTPLMGQAGTP 240  
DB 202 EQHDKQVCEREMORITPLSAFTNPTCEIVDEKTVVHTSQTVPDQEGSTPLMGQAGTP 261  
QY 241 GA 242  
DB 262 GA 263  
RESULT 4  
O00318 PRELIMINARY; PRT; 264 AA.  
ID O00318  
AC O00318  
DT 01-OCT-2003 (TrEMBLrel. 25, Created)  
DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)

```

01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Hypothetical protein.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Breast tumor;
RX MEDLINE=22389257; PubMed=12477932;
RA Klausner R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altshul S.F., Zeeberg B., Buetow K.H., Schaefer C.P., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diachenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahy J., Helton E., Ketteman W., Madan A., Rodrigues S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
RA Krzywicki M.I., Skalska U., Smailus D.E., Schnerch A., Schein J.E.,
RA Jones S.J., Marra M.A.
RC "Generation and initial analysis of more than 15,000 full-length human
RT and mouse cDNA sequences.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Breast tumor;
RX Klausner R.
RA Submitted (AUG-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC055920; AAHS5920.1; -.
KW Hypothetical protein.
SQ SEQUENCE 264 AA; 28567 MW; 833EA578FEB763A4 CRC64;

Query Match 79.8%; Score 1028.5; DB 11; Length 264;
Best Local Similarity 78.6%; Pred. No. 5.9e-87;
Matches 191; Conservative 16; Mismatches 35; Indels 1; Gaps 17

QY 1 SGGCTFDNGHLYREDQTSAPFGLRCLNWLDAQSLGASAPVSGAGNHSYCRNPDEDPGRGF 60
Db 22 SGGCTFDNGHLYREDQTSAPFGLRCLNWLAAQGSRESLTPSPGNHNYCRNPDPGRGF 81
QY 61 CVYSGEAGVPEKRCEDLRCPEITTSQA-LPAFTTETQESGPGADEVOVFAPANALPAR 119
Db 82 CVISSETGVPEKRCPEDCSETTSQAPPSSAMELEKSGAPGDKAEQVFPFANALPAR 141
QY 120 SEAAAVQPVIGISQKRVNRNSKEKKDLGTGLGVLTGMWIIIAIGAGIILGYSYKRGKDL 179
Db 142 SEAAEVQPVIGISQLVRMNSKEKKDLGTGLGVLTGMWIIIAIGAGIILGVYTKRGKDL 201
QY 180 KEQHOKVCEREMQRIITPLSAFNPICETVDEKTVVVHTSQTVPDPQSGSTPLMGQAGT 239
Db 202 KEQHKKACEREMQRIITPLSAFTNPCTCTVDENTIIIVHSNPADVQEGSTLLTGOAGT 261
QY 240 PGA 242
Db 262 PGA 264

RESULT 5
Q811Z2 PRELIMINARY; PRT; 264 AA.
ID Q811Z2
AC Q811Z2;
DT 01-JUN-2003 (TrEMBLrel. 24, Created)
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE HGFL(L) protein.
GN HGFL.

```

```

QY 1 SGGCFWNGHLYREDQTSAPGLRCLNWLDAOSGLASAPVSGAGNHSYCRNPDEDPGRPW 60
DB 22 SGGCFWNGHLYREDQTSAPGLRCLNWLDAOSGLASAPVSGAGNHSYCRNPDEDPGRPW 81
QY 61 CYYVSGEAGVPEKPCEDLRCPETTSQALPFTTEIOEAGSEGCADDEVQVAFANALPARS 120
DB 82 CYYVSGEAGVPEKPCEDLRCPETTSQALPFTTEIOEAGSEGCADDEVQVAFANALPARS 141
QY 121 EAAAVQPVIGISQVRVMSKCKDLGTLGVILGTMWIIIAIGAGIILGYKYR 175
DB 142 EAAAVQPVIGISQVRVMSKCKDLGTLGVILGTMWIIIAIGAGIILGYKYR 196

RESULT 7
Q81123 PRELIMINARY; PRT; 213 AA.
AC Q81123
DT 01-JUN-2003 (TrEMBLrel. 24, Created)
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DE HGFL(S) protein.
GN HGFL.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BALB/c;
RA Chiang H., Chang M.;
RL Submitted (JUL-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF528078; AAC33761.1; -.
DR InterPro; IPR000001; Kringle.
DR Pfam; PF00051; Kringle; 1.
DR PRINTS; PRO0018; KRINGLE.
DR SMART; SM00130; KR; 1.
DR PROSITE; PS00021; KRINGLE 1; 1.
DR PROSITE; PS00070; KRINGLE 2; 1.
SQ SEQUENCE 213 AA; 23011 MW; 4A5E3481025AE97 CRC64;

Query Match 57.0%; Score 734.5; DB 11; Length 213;
Best Local Similarity 77.8%; Pred. No. 7.3e-60;
Matches 137; Conservative 10; Mismatches 28; Indels 1; Gaps 1;

QY 1 SGGCFWNGHLYREDQTSAPGLRCLNWLDAOSGLASAPVSGAGNHSYCRNPDEDPGRPW 60
DB 22 SGGCFWNGHLYREDQTSAPGLRCLNWLDAOSGLASAPVSGAGNHSYCRNPDPGRPW 81
QY 61 CYYVSGEAGVPEKPCEDLRCPETTSQALPFTTEIOEAGSEGCADDEVQVAFANALPAR 119
DB 82 CYYISSETGVPEKPCEDVSCPETTSQAPPSSAMELEKSGAPGDKAEQVFPFANALPAR 141
QY 120 SEAAVQPVIGISQVRVMSKCKDLGTLGVILGTMWIIIAIGAGIILGYKYR 175
DB 142 SEAAVQPVIGISQVRVMSKCKDLGTLGVILGTMWIIIAIGAGIILGYKYR 197

RESULT 8
Q7SXB3 PRELIMINARY; PRT; 263 AA.
AC Q7SXB3
DT 01-OCT-2003 (TrEMBLrel. 25, Created)
DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DE Hypothetical protein.
OS Brachydanio rerio (Zebrafish) (Danio rerio).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;
OC Cyprinidae; Danio.
OX NCBI_TaxID=7955;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=AB; TISSUE=Body;

```

```

RX MEDLINE=22388257; PubMed=12477932;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buettow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Udén T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raba S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Halle S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Kettelman M., Madan A., Rodrigues S., Sanchez A.,
RA Whiting J., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
RA Krzywinski M.I., Skalska U., Smallos D.E., Schnerch A., Schein J.E.,
RA Jones S.J., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
RT and mouse cDNA sequences.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=AB; TISSUE=Body;
RA Strausberg R.;
RL Submitted (AUG-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC055675; AA055675.1; -.
KW Hypothetical protein.
SQ SEQUENCE 263 AA; 28777 MW; 8BEBCL17EC7C8A58 CRC64;

Query Match 47.4%; Score 611; DB 13; Length 263;
Best Local Similarity 47.9%; Pred. No. 2.6e-49;
Matches 125; Conservative 36; Mismatches 56; Indels 44; Gaps 7;

QY 4 CFWNGLHYREDQTSAPGLRCLNW-----LDAOSGLASAPVSGAGNHSYCRNPDEDP 56
DB 25 CITNNGEDYRGTOQTKTSSGTCLSWSRLNLFKFDQSQTGV-----GDHNFCKNPGSN 76
QY 57 RGPWCYSGEAGVPEKPCEDLRCPETTSQALPFTTEIOEAGSEGCADDEVQVAFANALPAR 106
DB 77 K-PWCYSGSGGKTKKACDIRICQD-----QNATEAPAPESVPTQGLTQR 122
QY 107 -VQVFPANALPARSEAAAVQPVIGISQVRVMSKCKDLGTLGVILGTMWIIIAIGA 165
DB 123 MVETEPANSPFSQVEGAAVQVKGVRQVRSRSGKCKDLGTLGVILAVFNMWIIILGG 182
QY 166 GILGYSYKRGKDLKEQHDQKVEREMORITPLSAFTNPTCEIVDEKTVVV---HTSQT 222
DB 183 GITMGYFYKRGDLKKQHEQRYVEREMHRTITPLSAFANFICELVDENTIVITAEPNQT 242
QY 223 PV-DPEGSTPLMGQAGTPGA 242
DB 243 PTQEPVEGADPLMGSAAGTPGA 263

RESULT 9
Q8SQ23 PRELIMINARY; PRT; 562 AA.
AC Q8SQ23
DT 01-JUN-2002 (TrEMBLrel. 21, Created)
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE T-plasminogen activator.
OS Sus scrofa (Pig).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
OX NCBI_TaxID=9823;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Enamel organ;
RA Ding Y., Xue J., Bartlett J.D.;
RT "T-plasminogen activator in tooth tissues.";
RL Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.

```

```
CC -!- SIMILARITY: CONTAINS 2 KRINGLE DOMAINS.
DR EMBL; AF364605; AAM00297.1; -.
DR HSSP; P00761; IAN1.
DR GO; GO:0005576; C:extracellular; IEA.
DR GO; GO:0004263; F:chymotrypsin activity; IEA.
DR GO; GO:0008233; F:peptidase activity; IEA.
DR GO; GO:0004235; F:trypsin activity; IEA.
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR InterPro; IPR009003; Cys_Ser_trypsin.
DR InterPro; IPR006209; EGF-like.
DR InterPro; IPR000083; Fibronctn1.
DR InterPro; IPR006210; IEGF.
DR InterPro; IPR000001; Kringle.
DR InterPro; IPR001254; Peptidase S1.
DR InterPro; IPR001314; Peptidase_S1A.
DR Pfam; PF00008; EGF; 1.
DR Pfam; PF00039; fnl; 1.
DR Pfam; PF00051; kringle; 2.
DR Pfam; PF00089; trypsin; 1.
DR PRINTS; PR00722; CHYMOTRYPSIN.
DR PRINTS; PR00018; KRINGLE.
DR ProDom; PD000395; Kringle; 2.
DR SMART; SM00181; EGF; 1.
DR SMART; SM00058; FNI; 1.
DR SMART; SM00130; KR; 2.
DR SMART; SM00020; Tryp_SPC; 1.
DR PROSITE; PS00022; EGF_1; 1.
DR PROSITE; PS01186; EGF_2; 1.
DR PROSITE; PS01253; FIBRONECTIN_1; 1.
DR PROSITE; PS00021; KRINGLE_1; 2.
DR PROSITE; PS00070; KRINGLE_2; 2.
DR PROSITE; PS00240; TRYPSIN_DOM; 1.
DR PROSITE; PS00134; TRYPSIN_HIS; 1.
DR PROSITE; PS00135; TRYPSIN_SER; 1.
DR EGF-like domain; Glycoprotein; Hydrolase; Kringle; Protease;
KW Serine protease.
SQ SEQUENCE 562 AA; 63668 MW; F986B4C77CB101E8 CRC64;

Query Match 13.0%; Score 167; DB 6; Length 562;
Best Local Similarity 36.6%; Pred. No. 8.8e-07;
Matches 41; Conservative 12; Mismatches 33; Indels 26; Gaps 6;

QY 4 CFWDNHLYREDQTSAPGLRCLNWLDAQSGLASAPVS-----GAGNHSYCRNPDE 54
Db 127 CYEDQGISYRGTWSTAESGAECTNW--NTSGLASHPYNGRRDPAVKLGLGNHNYCRNPDK 184

QY 55 DPRGPWCYV-SGAGVPEKPCEDLRCPETTSQALPAFTTRIQEASEGEGAD 105
Db 185 DSK-PWCYVFKAEKSPD-----PC-----STPACTKEKEECVTKGGLD 222

RESULT 10
Q727N2 PRELIMINARY; PRT; 291 AA.
AC Q727N2;
DT 01-OCT-2003 (T-EMBLrel. 25, Created)
DT 01-OCT-2003 (T-EMBLrel. 25, Last sequence update)
DT 01-OCT-2003 (T-EMBLrel. 25, Last annotation update)
DE Plasmalogen activator, tissue type isoform 2.
GN PLAT.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Rieder M.J., Carrington D.P., da Ponte S.H., Hastings N.C.,
RA Ahearn M.O., Kuldaneek S.A., Rajkumar N., Toth E.J., Yi Q.,
RA Nickerson D.A.;
RL Submitted (MAY-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY291060; AAP34246.1; -.
SQ SEQUENCE 291 AA; 32191 MW; 874E38C52F50EFID CRC64;
```

```
Query Match 12.0%; Score 154.5; DB 4; Length 291;
Best Local Similarity 39.6%; Pred. No. 5.4e-06;
Matches 36; Conservative 6; Mismatches 36; Indels 13; Gaps 4;

QY 4 CFWDNHLYREDQTSAPGLRCLNWLDAQSGLASAPVS-----GAGNHSYCRNPDE 54
Db 127 CYEDQGISYRGTWSTAESGAECTNW--NSSALAKPYSGRRPDARLGLGNHNYCRNPDR 184

QY 55 DPRGPWCYVSGAGVPEKPCEDLRCPETTS 85
Db 185 DSK-PWCYV-F-KAGKYSSEFCSTPACSEGS 213

RESULT 11
Q9BU99 PRELIMINARY; PRT; 516 AA.
AC Q9BU99;
DT 01-JUN-2001 (T-EMBLrel. 17, Created)
DT 01-JUN-2001 (T-EMBLrel. 17, Last sequence update)
DT 01-OCT-2003 (T-EMBLrel. 25, Last annotation update)
DE Similar to plasminogen activator, tissue.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Strausberg R.;
RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1.
CC -!- SIMILARITY: CONTAINS 2 KRINGLE DOMAINS.
DR EMBL; BC002795; AA02795.1; -.
DR HSP; P00750; IASH.
DR GO; GO:0004263; F:chymotrypsin activity; IEA.
DR GO; GO:0008233; F:peptidase activity; IEA.
DR GO; GO:0004295; F:trypsin activity; IEA.
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR InterPro; IPR009003; Cys_Ser_trypsin.
DR InterPro; IPR006209; EGF-like.
DR InterPro; IPR006210; IEGF.
DR InterPro; IPR000001; Kringle.
DR InterPro; IPR001254; Peptidase S1.
DR InterPro; IPR001314; Peptidase_S1A.
DR Pfam; PF00008; EGF; 1.
DR Pfam; PF00051; kringle; 2.
DR Pfam; PF00089; trypsin; 1.
DR PRINTS; PR00722; CHYMOTRYPSIN.
DR PRINTS; PR00018; KRINGLE.
DR ProDom; PD000395; Kringle; 2.
DR SMART; SM00181; EGF; 1.
DR SMART; SM00130; KR; 2.
DR SMART; SM00020; Tryp_SPC; 1.
DR PROSITE; PS00022; EGF_1; 1.
DR PROSITE; PS01186; EGF_2; 1.
DR PROSITE; PS00021; KRINGLE_1; 2.
DR PROSITE; PS00070; KRINGLE_2; 2.
DR PROSITE; PS00240; TRYPSIN_DOM; 1.
DR PROSITE; PS00134; TRYPSIN_HIS; 1.
DR PROSITE; PS00135; TRYPSIN_SER; 1.
DR EGF-like domain; Glycoprotein; Hydrolase; Kringle; Protease;
KW Serine protease.
SQ SEQUENCE 516 AA; 57370 MW; BAB31901FDC96800 CRC64;

Query Match 12.0%; Score 154.5; DB 4; Length 516;
Best Local Similarity 39.6%; Pred. No. 1.1e-05;
Matches 36; Conservative 6; Mismatches 36; Indels 13; Gaps 4;

QY 4 CFWDNHLYREDQTSAPGLRCLNWLDAQSGLASAPVS-----GAGNHSYCRNPDE 54
Db 81 CYEDQGISYRGTWSTAESGAECTNW--NSSALAKPYSGRRPDARLGLGNHNYCRNPDR 138

QY 55 DPRGPWCYVSGAGVPEKPCEDLRCPETTS 85
```



DB		139 DSK-PWCYVF-KAGKYSSEFCSTACSGNS 167	:	
RESULT 12				
Q86YK8	PRELIMINARY;	PRT; 562 AA.		
ID	Q86YK8			
AC	Q86YK8			
DT	01-JUN-2003 (TREMBLrel. 24, Created)			
DT	01-JUN-2003 (TREMBLrel. 24, Last sequence update)			
DT	01-OCT-2003 (TREMBLrel. 25, Last annotation update)			
DE	Tissue plasminogen activator.			
OS	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			
OX	NCBI_TaxID=9606;			
RN	[1]			
SEQUENCE FROM N.A.				
RA	Liu Y., Xu L., Zeng Y., He X.;			
RL	"cDNA of tissue plasminogen activator.";			
RT	Submitted (JAN-2003) to the EMBL/GenBank/DBJ databases.			
DR	EMBL; AY221101.1; AAC34406.1; -			
DR	GO; GO:0005576; C:extracellular; IEA.			
DR	GO; GO:0004263; F:chymotrypsin activity; IEA.			
DR	GO; GO:0004295; P:proteolysis and peptidolysis; IEA.			
DR	GO; GO:0006508; P:proteolysis and peptidolysis; IEA.			
DR	InterPro; IPR009003; Cys_Ser_trypsin.			
DR	InterPro; IPR006209; EGF-like.			
DR	Fam; PF00039; fnl; 1.			
DR	Fam; PF00051; kringle; 2.			
DR	Fam; PF00089; trypsin; 1.			
DR	PRINTS; PR00722; CHYMOTRYPSIN.			
DR	PRINTS; PR00018; KRINGLE.			
DR	ProDom; PD000395; Kringle; 2.			
DR	SMART; SM00058; FN1; 1.			
DR	SMART; SM00130; KR; 2.			
DR	SMART; SM00020; Tryp_Spc; 1.			
DR	PROSITE; PS00022; EGF_1; 2.			
DR	PROSITE; PS01186; EGF_2; 1.			
DR	PROSITE; PS01253; FIBRONECTIN_1; 1.			
DR	PROSITE; PS00021; KRINGLE_1; 1.			
DR	PROSITE; PS00021; KRINGLE_2; 1.			
DR	PROSITE; PS00070; KRINGLE_2; 1.			
DR	PROSITE; PS0240; TRYPSIN_DOM; 1.			
DR	PROSITE; PS00134; TRYPSIN_HIS; 1.			
DR	PROSITE; PS00135; TRYPSIN_SER; 1.			
KW	Hypothetical protein; EGF-like domain; Glycoprotein; Hydrolase;			
KW	Kringle; Protease; Serine protease			
SQ	SEQUENCE 562 AA; 62902 MW; 837D98392F6EDDF1 CRC64;			
Query Match	12.0%; Score 154.5; DB 4; Length 562;			
Best Local Similarity	39.6%; Pred. No. 1.3e-05;			
Matches	36; Conservative 6; Mismatches 36; Indels 13; Gaps 4;			
QY	4 CFWDNGHLYREDQTSAPGLRCNLWLDAOSGLASAPVS-----GAGNHSYCRNPDE 54			
DB	127 CYEQQSISYGTWTSAESGAECTNW--NSALAKPYSGRRPDATRLGLGNHYCRNPDR 184			
QY	55 DPRGPWCYVSGEAGVPKEPCEDLRCPETTS 85			
DB	185 DSK-PWCYVF-KAGKYSSEFCSTPACSGNS 213			
RESULT 13				
Q8VC84	PRELIMINARY;	PRT; 653 AA.		
ID	Q8VC84			
AC	Q8VC84			
DT	01-MAR-2002 (TREMBLrel. 20, Created)			
DT	01-MAR-2002 (TREMBLrel. 20, Last sequence update)			
DT	01-OCT-2003 (TREMBLrel. 25, Last annotation update)			
DE	Hypothetical protein.			

```
ID OSMB1 PRELIMINARY; PRT; 564 AA.
AC Q8MB1,
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Tissue-type plasminogen activator.
OS Oryctolagus cuniculus (Rabbit).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
OX NCBI_TaxID=9986;
RN [1]
RP SEQUENCE FROM N.A.
RA Sugiki M., Yoshida E., Anai K., Maruyama M.;
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1.
CC -!- SIMILARITY: CONTAINS 5 KRINGLE DOMAINS.
DR EMBL; AY029518; AAK40240.1; -.
DR GO; GO:0005509; F:calcium ion binding; IEA.
DR GO; GO:0004263; F:chymotrypsin activity; IEA.
DR GO; GO:0008233; F:peptidase activity; IEA.
DR GO; GO:0003809; F:thrombin activity; IEA.
DR GO; GO:0004295; F:trypsin activity; IEA.
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR InterPro; IPR009003; Cys Ser trypsin.
DR InterPro; IPR006209; EGF-like.
DR InterPro; IPR000083; Fibrinctn1.
DR InterPro; IPR006210; IEGF.
DR InterPro; IPR000001; Kringle.
DR InterPro; IPR001254; Peptidase_S1.
DR InterPro; IPR001314; Peptidase_S1A.
DR Pfam; PF00008; EGF; 1.
DR Pfam; PF00039; fni; 1.
DR Pfam; PF00051; kringle; 2.
DR Pfam; PF00089; trypsin; 1.
DR PRINTS; PR00722; CHYMOTRYPSIN.
DR PRINTS; PR00018; KRINGLE.
DR ProDom; PD000395; Kringle; 2.
DR SMART; SM00181; EGF; 1.
DR SMART; SM00058; fni; 1.
DR SMART; SM00130; Kr; 2.
DR SMART; SM00020; Tryp_Spc; 1.
DR PROSITE; PS00022; EGF_1; 1.
DR PROSITE; PS01186; EGF_2; 1.
DR PROSITE; PS01253; FIBRONECTIN_1; 1.
DR PROSITE; PS00021; KRINGLE_1; 2.
DR PROSITE; PS00070; KRINGLE_2; 2.
DR PROSITE; PS00240; TRYPSIN_DOM; 1.
DR PROSITE; PS00134; TRYPSIN_HIS; 1.
DR PROSITE; PS00135; TRYPSIN_SER; 1.
DR EGF-like domain; Glycoprotein; Hydrolase; Protease;
KW Serine protease.
SQ SEQUENCE 564 AA; 62726 MW; 459DBAC6D4A937C CRC64;

Query Match 11.3%; Score 145.5; DB 6; Length 564;
Best Local Similarity 39.5%; Pred. No. 8.6e-05;
Matches 34; Conservative 7; Mismatches 32; Indels 13; Gaps 4;

QY 4 CFWNGLHYREDQTSAPGLRCLNWLDAQSGLASAPVS-----GAGNHSYCRNPDE 54
Db 128 CYEDRGIGYGTWTSTESGACQVNW--NSSWLKPKYSGRKPVALRLGLGNHNYCRNPDR 185
QY 55 DPGFPCVYSGEAGVPEKPCEDLRC 80
Db 186 DTK-PWCYVF-RAGTYSPEFCSTPAC 209

RESULT 15
Q15146 PRELIMINARY; PRT; 810 AA.
AC Q15146;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Plasminogen precursor.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Browne M.J., Chapman C.G., Dodd I., Carey J.E., Lawrence G.M.P.,
RA Mitchell D., Robinson J.H.;
RT "Expression of recombinant human plasminogen and aglycoplasminogen in
RT HeLa cells";
RL Fibrinolysis 0:0-0(1991).
CC -!- SIMILARITY: CONTAINS 5 KRINGLE DOMAINS.
DR EMBL; M74220; AAA36451.1; -.
DR HSP; P00747; 2PK4.
DR GO; GO:0005509; F:calcium ion binding; IEA.
DR GO; GO:0004263; F:chymotrypsin activity; IEA.
DR GO; GO:0008233; F:peptidase activity; IEA.
DR GO; GO:0003809; F:thrombin activity; IEA.
DR GO; GO:0004295; F:trypsin activity; IEA.
DR GO; GO:0007596; P:blood coagulation; IEA.
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR InterPro; IPR009003; Cys Ser trypsin.
DR InterPro; IPR000001; Kringle.
DR InterPro; IPR003014; PAN.
DR InterPro; IPR003603; Pan_app.
DR InterPro; IPR001254; Peptidase_S1.
DR InterPro; IPR001314; Peptidase_S1A.
DR InterPro; IPR003966; Peptidase_S1A_pr.
DR Pfam; PF00051; kringle; 5.
DR Pfam; PF00024; PAN; 1.
DR Pfam; PF00089; trypsin; 1.
DR PRINTS; PR00722; CHYMOTRYPSIN.
DR PRINTS; PR00018; KRINGLE.
DR ProDom; PD000395; PROTHROMBIN.
DR SMART; SM00130; KR; 5.
DR SMART; SM00473; PAN AP; 1.
DR SMART; SM00020; Tryp_Spc; 1.
DR PROSITE; PS00021; KRINGLE_1; 5.
DR PROSITE; PS00070; KRINGLE_2; 5.
DR PROSITE; PS00240; TRYPSIN_DOM; 1.
DR PROSITE; PS00134; TRYPSIN_HIS; 1.
DR PROSITE; PS00135; TRYPSIN_SER; 1.
DR Glycoprotein; Hydrolase; Kringle; Protease; Serine protease; Signal.
KW SIGNAL 1 19 POTENTIAL.
FT CHAIN 20 810 PLASMINOGEN.
SQ SEQUENCE 810 AA; 90555 MW; B05C7D4B0D020B3C CRC64;

Query Match 11.0%; Score 142; DB 4; Length 810;
Best Local Similarity 37.2%; Pred. No. 0.00029;
Matches 32; Conservative 8; Mismatches 34; Indels 12; Gaps 4;

QY 4 CFWNGLHYREDQTSAPGLRCLNWLDA-----QSGLASAPVSGAGNHSYCRNPDEPRG 58
Db 103 CKTNGKRYRTMSKXNGITCKXWSSTSPHRPFSFATHPEGL-EENYCRNPDNPDQG 161
QY 59 PWCYVSGEAGVPEKPCEDLRCPE 82
Db 162 PWCYTTD-----PKRYDYCDILECEE 183

Search completed: March 17, 2004, 07:06:55
Job time : 44.2961 secs
```

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 17, 2004, 06:55:13 ; Search time 61.8843 Seconds  
(without alignments)  
1104.909 Million cell updates/sec

Title: US-10-057-951-2\_COPY\_22\_263  
Perfect score: 1289  
Sequence: 1 SGCCFWNGLHLYREDQTSFA.....PVDQEGSTPLMGQAGTGA 242

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A\_Geneseq\_29Jan04:\*

- 1: Geneseqp1980s:\*
- 2: Geneseqp1990s:\*
- 3: Geneseqp2000s:\*
- 4: Geneseqp2001s:\*
- 5: Geneseqp2002s:\*
- 6: Geneseqp2003as:\*
- 7: Geneseqp2003bs:\*
- 8: Geneseqp2004s:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1289	100.0	263	2 AAW87769	AAW87769 Human tis
2	1289	100.0	263	2 AAY05219	AAY05219 Kringle1
3	1289	100.0	263	4 AAE00300	AAE00300 Human tis
4	1289	100.0	263	5 ABR40414	ABR40414 Human sec
5	1286	99.8	263	5 AAU86149	AAU86149 Human PRO
6	1280	99.3	263	3 AAB43237	AAB43237 Human ORF
7	1275	98.9	263	4 AAM93748	AAM93748 Human pol
8	1063.5	82.5	286	2 AAY05220	AAY05220 Kringle1
9	648	50.3	146	5 ABR40487	ABR40487 Human sec
10	648	50.3	146	5 ABR40561	ABR40561 Human sec
11	385	29.9	81	7 ABR42624	ABR42624 Human kri
12	322	25.0	66	4 AAM18800	AAM18800 Peptide #
13	322	25.0	66	4 ABB37905	ABB37905 Peptide #
14	322	25.0	66	4 AAM31314	AAM31314 Peptide #
15	322	25.0	66	4 ABB23159	ABB23159 Protein #
16	322	25.0	66	4 AAM71037	AAM71037 Human bon
17	322	25.0	66	4 AAM58537	AAM58537 Human bra
18	322	25.0	66	4 ABG52752	ABG52752 Human liv
19	322	25.0	66	5 ABG40828	ABG40828 Human pep
20	202	15.7	56	2 AAY12615	AAY12615 Human 5'
21	201	15.6	55	2 AAY12397	AAY12397 Human 5'
22	196	15.2	39	2 AAW72641	AAW72641 Nervous g
23	192	14.9	39	2 AAW72640	AAW72640 Nervous g
24	160.5	12.5	527	2 AAR20220	AAR20220 t-PA anal
25	160.5	12.5	527	2 AAR20219	AAR20219 t-PA anal

26	160.5	12.5	527	2 AAR20217	AAR20217 t-PA anal
27	160.5	12.5	527	2 AAR20218	AAR20218 t-PA anal
28	159.5	12.4	527	2 AAR20221	AAR20221 t-PA anal
29	159.5	12.4	527	2 AAR20222	AAR20222 t-PA anal
30	159.5	12.4	527	2 AAR20223	AAR20223 t-PA anal
31	157.5	12.2	527	2 AAW54157	AAW54157 t-PA muta
32	157.5	12.2	527	2 AAW54154	AAW54154 t-PA muta
33	156.5	12.1	439	2 AAR68851	AAR68851 Delta 2-8
34	156.5	12.1	483	2 AAR70879	AAR70879 Human tis
35	156.5	12.1	483	2 AAR70883	AAR70883 Human tis
36	156.5	12.1	483	2 AAR70884	AAR70884 Human tis
37	156.5	12.1	483	2 AAR70886	AAR70886 Human tis
38	156.5	12.1	483	2 AAR70878	AAR70878 Human tis
39	156.5	12.1	483	2 AAR70877	AAR70877 Human tis
40	156.5	12.1	483	2 AAR70887	AAR70887 Human tis
41	156.5	12.1	483	2 AAR70881	AAR70881 Human tis
42	156.5	12.1	483	2 AAR70882	AAR70882 Human tis
43	156.5	12.1	483	2 AAR70885	AAR70885 Human tis
44	156.5	12.1	483	2 AAR70889	AAR70889 Human tis
45	156.5	12.1	483	2 AAR70888	AAR70888 Human tis

## ALIGNMENTS

RESULT 1

AAW87769

ID AAW87769 standard; protein; 263 AA.

AC AAW87769;

DT 29-MAR-1999 (first entry)

DE Human tissue plasminogen activator-like protease t-PALP.

KW Tissue plasminogen activator-like protease; t-PALP; human;  
KW circulatory system-related disorder; blood clotting; stroke; thrombosis;  
KW peripheral arterial occlusion; pulmonary embolism; myocardial infarction;  
KW diagnosis; therapy.

OS Homo sapiens.

FX Key Location/Qualifiers

FT Peptide 1..21 /label= Sig\_peptide

FT Protein 22..263 /label= Mat\_protein

FT Peptide 22..31 /note= "epitope-bearing region"

FT Domain 25..84 /note= "kringle domain"

FT Peptide 35..44 /note= "epitope-bearing region"

FT Peptide 71..81 /note= "epitope-bearing region"

FT Domain 85..263 /note= "protease domain"

FT Peptide 91..107 /note= "epitope-bearing region"

FT Peptide 119..128 /note= "epitope-bearing region"

FT Peptide 138..147 /note= "epitope-bearing region"

FT Peptide 155..167 /note= "epitope-bearing region"

FT Peptide 193..203 /note= "epitope-bearing region"

FT Peptide 206..215 /note= "epitope-bearing region"

FT Peptide 227..237 /note= "epitope-bearing region"

FT Peptide 243..252 /note= "epitope-bearing region"

FT Peptide 243..252 /note= "epitope-bearing region"

XX PN WO9854199-A1.  
 XX XX 03-DEC-1998.  
 XX XX 27-MAY-1998; 98WO-US010728.  
 XX XX 28-MAY-1997; 97US-0048000P.  
 XX XX (HUMA-) HUMAN GENOME SCI INC.  
 XX XX Ebner R, Moore PA, Ruben SM;  
 XX XX WPI; 1999-070207/06.  
 XX XX N-PSDB; AA999636.  
 XX XX New tissue plasminogen activator-like protease - useful in the diagnosis  
 XX XX and treatment of circulatory system-related disorders.  
 XX XX Claim 1; Page 56-57; 76pp; English.  
 XX XX This is the amino acid sequence of tissue plasminogen activator-like  
 XX XX protease (t-PALP), a novel member of the serine protease family that  
 XX XX shares sequence homology to human tissue plasminogen activator (see  
 XX XX AA87770). The t-PALP sequence was deduced from a cDNA clone (see  
 XX XX AA999636) derived from activated monocytes. The 2.5 kb t-PALP message has  
 XX XX also been detected in heart, brain, lung, placenta, liver, skeletal  
 XX XX muscle, kidney, pancreas, spleen, thymus, prostate, testis, ovary, small  
 XX XX intestine, colon and peripheral blood leukocytes. Isolated nucleic acids  
 XX XX encoding amino acids 21 to 242, 20 to 242, 1-242, 4-63 (kringle domain)  
 XX XX and 84-242 (protease domain) of t-PALP, or encoding epitope-bearing  
 XX XX portions of t-PALP, are also claimed, as are recombinant vectors, host  
 XX XX cells, and methods for producing t-PALP polypeptides. t-PALP may be used  
 XX XX to detect and treat disorders related to the circulatory system, and to  
 XX XX identify agonists and antagonists of t-PALP activity. The homology  
 XX XX between t-PALP and tPA indicates that t-PALP may be involved in the  
 XX XX regulation of normal and abnormal clotting in e.g. stroke, deep-vein  
 XX XX thrombosis, peripheral arterial occlusion, pulmonary embolism and  
 XX XX myocardiothrombosis  
 XX XX Sequence 263 AA;  
 XX XX  
 XX Query Match 100.0%; Score 1289; DB 2; Length 263;  
 XX Best Local Similarity 100.0%; Pred. No. 1.5e-113;  
 XX Matches 242; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 SGGCFWNGHLYREDQTSAPGLRCLNWLDAQSLASAPVSGAGNHSYCRNPDEPRGPW 60  
 Db 22 SGGCFWNGHLYREDQTSAPGLRCLNWLDAQSLASAPVSGAGNHSYCRNPDEPRGPW 81  
 QY 61 CVVSGEAGVPEKPCEDLRCPETTSQALPAFTTEIQEASGPGADVQVFPANALPARS 120  
 Db 82 CVVSGEAGVPEKPCEDLRCPETTSQALPAFTTEIQEASGPGADVQVFPANALPARS 141  
 QY 121 EAAAVQPVIGISQVRVNSKEKDLGTLGYVLGITMMVIIIAIGAGIILGYSGYRGKDLK 180  
 Db 142 EAAAVQPVIGISQVRVNSKEKDLGTLGYVLGITMMVIIIAIGAGIILGYSGYRGKDLK 201  
 QY 181 EQHDKVCEMORITLPLSAFTNPTCEIVDEKTVVHTSOTVPDQEGSTPLMGOAGTP 240  
 Db 202 EQHDKVCEMORITLPLSAFTNPTCEIVDEKTVVHTSOTVPDQEGSTPLMGOAGTP 261  
 QY 241 GA 242  
 Db 262 GA 263

RESULT 2  
 ID AA905219  
 XX AA905219 standard; protein; 263 AA.  
 AC AA905219;  
 XX

DT 17-JUN-1999 (first entry)  
 XX Kringle1 protein sequence.  
 XX Kringle1; cancer; inflammation; autoimmunity; rheumatoid arthritis; AIDS;  
 XX CNS inflammation; cerebellar degeneration; Alzheimer's disease; asthma;  
 XX Parkinson's disease; multiple sclerosis; amyotrophic lateral sclerosis;  
 XX neurological abnormality; ischaemia reperfusion injury; ischaemic injury;  
 XX cardiovascular disease; kidney disease; liver disease; aplastic anaemia;  
 XX myocardial infarction; hypotension; hypertension; allergy; infection;  
 XX myelodysplastic syndrome; haematologic abnormality; diagnosis; therapy;  
 XX male pattern baldness.  
 OS Homo sapiens.  
 XX WO9911788-A1.  
 XX 11-MAR-1999.  
 XX 02-SEP-1998; 98WO-US018270.  
 XX 02-SEP-1997; 97US-0056032P.  
 XX 01-SEP-1998; 98US-00144889.  
 XX (SMIK ) SMITHKLINE BEECHAM CORP.  
 XX Albone EF, Kikly KK;  
 XX WPI; 1999-214707/18.  
 XX N-PSDB; AAX28354.  
 XX New kringle1 polypeptides and polynucleotides.  
 XX Claim 1; Page 31-32; 42pp; English.  
 XX This sequence is a Kringle1 polypeptide of the invention. The kringle1  
 XX polypeptides (I) are used to screen for agonists and antagonists.  
 XX Agonists are used to treat subjects in need of enhanced activity or  
 XX expression of (I). Antagonists are used to treat subjects having need to  
 XX inhibit the activity or expression of (I). The methods can be used to  
 XX treat conditions such as cancer, inflammation, autoimmunity, allergy,  
 XX asthma, rheumatoid arthritis, CNS inflammation, cerebellar degeneration,  
 XX Alzheimer's disease, Parkinson's disease, multiple sclerosis,  
 XX amyotrophic lateral sclerosis, head injury damage and other neurological  
 XX abnormalities, ischaemia reperfusion injury, cardiovascular disease,  
 XX kidney disease, liver disease, ischaemic injury, myocardial infarction,  
 XX hypotension, hypertension, AIDS, myelodysplastic syndromes and other  
 XX haematologic abnormalities, aplastic anaemia, male pattern baldness, and  
 XX bacterial, fungal, protozoan and viral infections. The kringle1  
 XX polypeptides may also be used to generate antibodies. Determining the  
 XX presence or absence of mutations in, and analysing for the presence or  
 XX absence of expression of, kringle1 polynucleotides can be used to  
 XX diagnose a disease or susceptibility to a disease related to expression  
 XX or activity of kringle1 proteins. The polynucleotides may also be used  
 XX for chromosome identification, and mapping  
 XX Sequence 263 AA;  
 XX  
 XX Query Match 100.0%; Score 1289; DB 2; Length 263;  
 XX Best Local Similarity 100.0%; Pred. No. 1.5e-113;  
 XX Matches 242; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 SGGCFWNGHLYREDQTSAPGLRCLNWLDAQSLASAPVSGAGNHSYCRNPDEPRGPW 60  
 Db 22 SGGCFWNGHLYREDQTSAPGLRCLNWLDAQSLASAPVSGAGNHSYCRNPDEPRGPW 81  
 QY 61 CVVSGEAGVPEKPCEDLRCPETTSQALPAFTTEIQEASGPGADVQVFPANALPARS 120  
 Db 82 CVVSGEAGVPEKPCEDLRCPETTSQALPAFTTEIQEASGPGADVQVFPANALPARS 141  
 QY 121 EAAAVQPVIGISQVRVNSKEKDLGTLGYVLGITMMVIIIAIGAGIILGYSGYRGKDLK 180  
 Db 142 EAAAVQPVIGISQVRVNSKEKDLGTLGYVLGITMMVIIIAIGAGIILGYSGYRGKDLK 201



Matches 242; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 SGGCFWMDNGHLYREDQTSAPGLRCLNWLDAQSLASAPVSGAGNHSYCRNPDEDPGPW 60  
 Db 22 SGGCFWMDNGHLYREDQTSAPGLRCLNWLDAQSLASAPVSGAGNHSYCRNPDEDPGPW 81  
 QY 61 CVYSGEAGVPEKPCEDLRCPETTSQALPAFTTEIQEASEGPGADEVQVFPANALPARS 120  
 Db 82 CVYSGEAGVPEKPCEDLRCPETTSQALPAFTTEIQEASEGPGADEVQVFPANALPARS 141  
 QY 121 EAAAVQPVIGISQVRMNSKEKDLGTGLGYLGTVMWIIIAIGAGIILGYSYKRGKDLK 180  
 Db 142 EAAAVQPVIGISQVRMNSKEKDLGTGLGYLGTVMWIIIAIGAGIILGYSYKRGKDLK 201  
 QY 181 EQHDKVCEREMQRIITPLSAFTNPTCEIVDEKTVVHTSQTVPDPOEGSTPLMGQAGTP 240  
 Db 202 EQHDKVCEREMQRIITPLSAFTNPTCEIVDEKTVVHTSQTVPDPOEGSTPLMGQAGTP 261  
 QY 241 GA 242  
 Db 262 GA 263

RESULT 4  
 ABR40414  
 ID ABR40414 standard; protein; 263 AA.

XX AC ABR40414;  
 XX XX  
 DT 13-JUN-2003 (first entry)  
 XX XX  
 DE Human secreted protein #SEQ ID 164.  
 XX XX  
 KW Human; secreted protein; anti-HIV; neutropic; neuroprotective;  
 KW antianigmal; immunosuppressive; immunomodulator; cytostatic; cardiant;  
 KW hepatotropic; antiinflammatory; antiallergic; antidiabetic;  
 KW gastrointestinal; infertility; nephrotropic; virucide; hypotensive;  
 KW vasotropic; dermatological; osteopathic; antiarthritic; antiparkinsonian;  
 KW antiaesthetic; antipsoriatic; cerebroprotective; antibacterial;  
 KW fungicide; antiparasitic; protein therapy; gene therapy; cancer;  
 KW hyperproliferative disorder; leukaemia; autoimmune disorder;  
 KW immunodeficiency; HIV; infertility; uterine disorders; endometriosis;  
 KW amenorrhea; ocular disorder; neurological disorder; wound healing;  
 KW Huntington's disease; gastrointestinal disorder; inflammatory disease;  
 KW Crohn's disease; kidney disorder; renal failure; diabetes; obesity;  
 KW cirrhosis; hepatitis; cardiovascular disorder; angina; osteoarthritis;  
 KW rheumatoid arthritis; psoriasis; multiple sclerosis; asthma;  
 KW respiratory disorder; infectious disease; chromosome identification;  
 KW food additive; nutrition.

XX OS Homo sapiens.  
 XX XX  
 XX WO200268628-A1.  
 XX XX  
 PD 06-SEP-2002.  
 XX XX  
 PF 21-FEB-2002; 2002WO-US005301.  
 XX XX  
 PR 23-FEB-2001; 2001US-0270625P.  
 PR 12-JUL-2001; 2001US-0304417P.  
 XX XX  
 PA (HUMA-) HUMAN GENOME SCI INC.  
 XX XX  
 XX Rosen CA, Komatsoulis G, Baker KP, Fiscella M, Moore PA, Wei P;  
 FI Duan DR, Shi Y, Gupta R;  
 XX XX  
 DR WPI; 2002-750417/81.  
 DR N-PSDB; ABZ82469.  
 XX XX  
 PT New human secreted proteins and nucleic acids, useful for preventing,  
 PT treating, or ameliorating a medical condition, e.g. cancer, autoimmune  
 PT disorder, HIV, Parkinson's disease, Alzheimer's disease, diabetes,  
 PT obesity or cirrhosis.

XX PS  
 CC The invention relates to novel human secreted proteins and the genes  
 CC encoding them. Genes and proteins of the invention may be useful for  
 CC preventing, treating or ameliorating medical conditions e.g. by protein  
 CC or gene therapy. These conditions include cancer and hyperproliferative  
 CC autoimmune disorders, immunodeficiencies (e.g. HIV infection),  
 CC disorders, immune cell proliferative disorders (e.g. leukaemia),  
 CC anemorrhea, ocular disorders, neurological disorders (e.g. Huntington's  
 CC disease), wound healing, gastrointestinal system disorders, particularly  
 CC inflammatory diseases (e.g. Crohn's disease), kidney disorders (e.g.  
 CC renal failure), diabetes, obesity, cirrhosis, hepatitis, cardiovascular  
 CC disorders (e.g. angina), rheumatoid arthritis, osteoarthritis, psoriasis,  
 CC multiple sclerosis, respiratory disorders (e.g. asthma), or infectious  
 CC diseases caused by bacterial, parasitic, viral or fungal agents. The  
 CC nucleic acids are also useful for chromosome identification, radiation  
 CC hybrid mapping or long-range restriction mapping. The polypeptide,  
 CC polynucleotide, agonist or antagonist may also be used as a food additive  
 CC or preservative to increase or decrease storage capabilities, fat content  
 CC or other nutritional components. The sequences given in records ABR40409-  
 CC ABR40590 and ABZ82464-ABZ82611 represent human secreted proteins and the  
 CC genes encoding them  
 XX SQ Sequence 263 AA;

Query Match 100.0%; Score 1289; DB 5; Length 263;  
 Best Local Similarity 100.0%; Pred No. 1.5e-113;  
 Matches 242; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 SGGCFWMDNGHLYREDQTSAPGLRCLNWLDAQSLASAPVSGAGNHSYCRNPDEDPGPW 60  
 Db 22 SGGCFWMDNGHLYREDQTSAPGLRCLNWLDAQSLASAPVSGAGNHSYCRNPDEDPGPW 81  
 QY 61 CVYSGEAGVPEKPCEDLRCPETTSQALPAFTTEIQEASEGPGADEVQVFPANALPARS 120  
 Db 82 CVYSGEAGVPEKPCEDLRCPETTSQALPAFTTEIQEASEGPGADEVQVFPANALPARS 141  
 QY 121 EAAAVQPVIGISQVRMNSKEKDLGTGLGYLGTVMWIIIAIGAGIILGYSYKRGKDLK 180  
 Db 142 EAAAVQPVIGISQVRMNSKEKDLGTGLGYLGTVMWIIIAIGAGIILGYSYKRGKDLK 201  
 QY 181 EQHDKVCEREMQRIITPLSAFTNPTCEIVDEKTVVHTSQTVPDPOEGSTPLMGQAGTP 240  
 Db 202 EQHDKVCEREMQRIITPLSAFTNPTCEIVDEKTVVHTSQTVPDPOEGSTPLMGQAGTP 261  
 QY 241 GA 242  
 Db 262 GA 263

RESULT 5  
 AAU86149  
 ID AAU86149 standard; protein; 263 AA.

XX AC AAU86149;  
 XX XX  
 DT 15-JUL-2002 (first entry)  
 XX XX  
 DE Human PRO264 polypeptide.  
 XX XX  
 KW Human; PRO; benign tumour; malignant tumour; lymphoid malignancy;  
 KW leukaemia; neuronal disorder; stromal disorder; blastocoele disorder;  
 KW inflammatory disorder; immune disorder; angiogenic disorder; cytostatic;  
 KW neuroprotective.  
 XX OS Homo sapiens.  
 XX XX  
 XX WO200153486-A1.  
 XX XX  
 PD 26-JUL-2001.

PF 11-FEB-2000; 2000WO-US003565.  
 XX 08-MAR-1999; 99WO-US005028.  
 PR 11-MAR-1999; 99US-0123972P.  
 PR 11-MAR-1999; 99US-0133459P.  
 PR 02-JUN-1999; 99WO-US012252.  
 PR 22-JUN-1999; 99US-0140650P.  
 PR 22-JUN-1999; 99US-0140653P.  
 PR 26-JUL-1999; 99US-0144758P.  
 PR 28-JUL-1999; 99US-0145698P.  
 PR 17-AUG-1999; 99US-0149395P.  
 PR 21-AUG-1999; 99US-0151689P.  
 PR 01-SEP-1999; 99WO-US020111.  
 PR 15-SEP-1999; 99WO-US021090.  
 PR 30-NOV-1999; 99WO-US028313.  
 PR 01-DEC-1999; 99WO-US028301.  
 PR 01-DEC-1999; 99WO-US028634.  
 PR 05-JAN-2000; 2000WO-US000219.  
 XX (GETH ) GENENTECH INC.  
 XX Ashkenazi AJ, Goddard A, Godowski PJ, Gurney AL, Hillan KJ;  
 PI Marsters SA, Pan J, Pitti RM, Roy MA, Smith V, Stone DM;  
 PI Watanabe CK, Wood WI;  
 XX WPI; 2002-205567/26.  
 DR N-PSDB; ABK40275.  
 XX  
 XX Thirty five nucleic acids encoding PRO polypeptides, useful for treating  
 PT benign or malignant tumors, leukemias and lymphoid malignancies,  
 PT inflammatory, angiogenic and immunologic disorders.  
 XX  
 XX Claim 61; Fig 44; 302pp; English.  
 XX  
 XX The present invention relates to the isolation of novel human PRO  
 CC polypeptides and the polynucleotide sequences encoding them. The PRO  
 CC polypeptides, agonists, antagonists or anti-PRO antibodies are useful for  
 CC treating benign or malignant tumors (e.g. renal, kidney, bladder,  
 CC breast, etc), leukemias and lymphoid malignancies, other disorders such  
 CC as neuronal, glial, astrocytal, hypothalamic, glandular, macrophagal,  
 CC stromal and blastocoeleic disorders, inflammatory, immune and angiogenic  
 CC disorders. The polynucleotide sequences are also useful in gene therapy.  
 CC AAU86128-AAU86162 represent the human PRO polypeptides of the invention  
 XX  
 XX Sequence 263 AA;  
 SQ

Query Match 99.8%; Score 1286; DB 5; Length 263;  
 Best Local Similarity 99.6%; Pred. No. 2.9e-113;  
 Matches 241; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 SGGCFWNGHLYREDQTSAPGLRCLNWLDAQSGLASAPVSGAGNHSYCRNPDDPRGPW 50  
 Db 22 SGGCFWNGHLYREDQTSAPGLRCLNWLDAQSGLASAPVSGAGNHSYCRNPDDPRGPW 81  
 Qy 61 CYVSGEAGVPEKRCEDLRCPETTSQALPAFTTEIQASGPGADEVQVFPANALPARS 120  
 Db 82 CYVSGEAGVPEKRCEDLRCPETTSQALPAFTTEIQASGPGADEVQVFPANALPARS 141  
 Qy 121 EAAAVQVIGISQVRVNSKEKDLGTLGYVLGTMVVIITAGIILGYSYKRGKDLK 180  
 Db 142 EAAAVQVIGISQVRVNSKEKDLGTLGYVLGTMVVIITAGIILGYSYKRGKDLK 201  
 Qy 181 EHQPKYCEMERQITPLSNFTNPTCEIYDEKTVVHTSQTVPDQEGSTPLMGAGTP 240  
 Db 202 EHQPKYCEMERQITPLSNFTNPTCEIYDEKTVVHTSQTVPDQEGSTPLMGAGTP 261  
 Qy 241 GA 242  
 Db 262 GA 263

RESULT 6

AA43237  
 ID AAB43237 standard; protein; 263 AA.  
 XX AAB43237;  
 AC  
 DT 08-FEB-2001 (first entry)  
 XX  
 DE Human ORFX ORF3001 polypeptide sequence SEQ ID NO:6002.  
 XX  
 KW Human; open reading frame; ORFX; detection; cytostatic; hepatotropic;  
 KW vulnery; antiprosiatic; antiparkinsonian; nootropic; neuroprotective;  
 KW anticonvulsant; osteopathic; antiarthritic; immunosuppressant; cardiant;  
 KW immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;  
 KW hypotensive; dermatological; immunosuppressive; antinflammatory;  
 KW antiviral; antibacterial; antifungal; antirheumatic; antithyroid;  
 KW antianemic; gene therapy; cancer; proliferative disorder; hypertension;  
 KW neurodegenerative disorder; osteoarthritis; graft vs host disease;  
 KW cardiovascular disease; diabetes mellitus; erythematous; SCID; AIDS;  
 KW cholesterol ester storage; systemic lupus erythematosus; infection;  
 KW severe combined immunodeficiency; malaria; autoimmune disorder; asthma;  
 KW allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;  
 KW bone damage; cartilage damage; antinflammatory disease; coagulation;  
 KW thrombosis; contraceptive.  
 XX Homo sapiens.  
 OS  
 XX WO200058473-A2.  
 XX  
 PD 05-OCT-2000.  
 XX  
 XX 31-MAR-2000; 2000WO-US008621.  
 PF  
 XX 31-MAR-1999; 99US-0127607P.  
 PR 02-APR-1999; 99US-0127636P.  
 PR 05-APR-1999; 99US-0127728P.  
 PR 30-MAR-2000; 2000US-00540763.  
 XX  
 XX (CURA-) CURAGEN CORP.  
 PA  
 XX Shimkets RA, Leach M;  
 PI  
 XX WPI; 2000-602362/57.  
 DR N-PSDB; AAC77446.  
 DR  
 XX Novel nucleic acids and peptides derived from open reading frame X,  
 PT useful for treating e.g. cancers, proliferative disorders,  
 PT neurodegenerative disorders and cardiovascular disease.  
 XX  
 XX Claim 11; Page 5181-5182; 5507pp; English.  
 PS  
 XX AAC7446 to AAC77606 encode the proteins given in AAB40237 to AAB43397,  
 CC which represent the human ORFX open reading frames 1 to 3161. The ORFX  
 CC sequences have activities such as: cytostatic; hepatotropic; vulnery;  
 CC antiprosiatic; antiparkinsonian; nootropic; neuroprotective; osteopathic;  
 CC anticonvulsant; antiarthritic; immunosuppressant; immunostimulant;  
 CC cardiant; thrombolytic; coagulant; vasotropic; antidiabetic; hypotensive;  
 CC dermatological; immunosuppressive; antinflammatory; antibacterial;  
 CC antiviral; antifungal; antirheumatic; antithyroid; antianemic. The  
 CC sequences can be used for determining the presence of or predisposition  
 CC to, or preventing or treating pathological conditions associated with an  
 CC ORFX-associated disorder. The nucleic acids can be used to express ORFX  
 CC proteins in gene therapy vectors. The proteins and nucleic acids may be  
 CC used to treat cancers, proliferative disorders, neurodegenerative  
 CC disorders, osteoarthritis, graft vs host disease, cardiovascular disease,  
 CC diabetes mellitus, hypertension, hypothyroidism, cholesterol ester  
 CC storage, systemic lupus erythematosus, severe combined immunodeficiency  
 CC (SCID), AIDS, viral, bacterial or fungal infection, malaria, autoimmune  
 CC disorders, asthma, allergies, aplastic anaemia, burns, wounds, bone and  
 CC cartilage damage, nocturnal haemoglobinuria, antinflammatory disease, to  
 CC enhance coagulation; to inhibit thrombosis; and as a contraceptive  
 XX  
 XX Sequence 263 AA;

Query Match 99.3%; Score 1280; DB 3; Length 263;  
 Best Local Similarity 99.2%; Pred. No. 1.1e-112;  
 Matches 240; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 SGGCFWNGHLYREDQTSAPGLRCLNWLDAQSLASAPVSGAGNHSYCRNPDEDPGPW 60  
 DB 22 SGGCFWNGHLYREDQTSAPGLRCLNWLDAQSLASAPVSGAGNHSYCRNPDEDPGPW 81

QY 61 CYYSGEAGVPEKPCEDLRCPETTSQALPAFTTEIQEASEGPGADVQVAFANALPARS 120  
 DB 82 CYYSGEAGVPEKPCEDLRCPETTSQALPAFTTEIQEASEGPGADVQVAFANALPARS 141

QY 121 EAAAVQPVIGISORVRMNSKEKDLGTGLGVLTGVTWVHIIAIGAGIILGYSYKRGKDLK 180  
 DB 142 EAAAVQPVIGISORVRMNSKEKDLGTGLGVLTGVTWVHIIAIGAGIILGYSYKRGKDLK 201

QY 181 EQHDQKVCEREMQRIITPLSAFTNPTCEIVDEKTVVHTSQTPVDPQEGSTPLMQAGTP 240  
 DB 202 EQHDQKVCEREMQRIITPLSAFTNPTCEIVDEKTVVHTSQTPVDPQEGSTPLMQAGTP 261

QY 241 GA 242  
 DB 262 GA 263

RESULT 7  
 AAM93748  
 ID AAM93748 standard; protein; 263 AA.  
 AC AAM93748;  
 XX  
 DT 06-NOV-2001 (first entry)  
 DE Human polypeptide, SEQ ID NO: 3727.  
 XX Human; full length cDNA; cDNA synthesis; oligo-capping.  
 KW Homo sapiens.  
 OS  
 XX  
 PN EP1130094-A2.  
 XX  
 PD 05-SEP-2001.  
 XX  
 PF 07-JUL-2000; 2000EP-00114089.  
 XX  
 PR 08-JUL-1999; 99JP-00194486.  
 PR 11-JAN-2000; 2000JP-00118774.  
 PR 02-MAY-2000; 2000JP-00183765.  
 XX  
 PA (HELI-) HELIX RES INST.  
 XX  
 PI Ota T, Nishikawa T, Isogai T, Hayaashi K, Ishii S, Kawai Y;  
 PI Wakamatsu A, Sugiyama T, Nagai K, Kojima S, Otsuki T, Koga H;  
 XX  
 DR WPI; 2001-524255/58.  
 DR N-PSDB; AAK94700.  
 XX  
 PT 830 Primers useful for synthesizing full length cDNA clones and their use  
 PT in genetic manipulation.  
 XX  
 PS Claim 8; SEQ ID NO 3727; 1380pp + Sequence Listing; English.  
 XX

The invention relates to primers for synthesizing full length cDNA clones. 830 cDNA molecules encoding a human protein have been isolated and nucleotide sequences of 5' and 3'-ends of the cDNA molecules have been determined. Primers for synthesizing the full length cDNA are useful for clarifying the function of the protein encoded by the cDNA. The full length clones were obtained by construction of full length enriched cDNA libraries that were synthesised by the oligo-capping method. The primers enable the production of the full length cDNA easily without any special methods. The present sequence is a polypeptide encoded by a full length human cDNA of the invention. Note: The sequence data for this patent did not form part of the printed specification, but was obtained in CD-ROM

CC format directly from BPO  
 XX  
 SQ Sequence 263 AA;

Query Match 98.9%; Score 1275; DB 4; Length 263;  
 Best Local Similarity 99.2%; Pred. No. 3.2e-112;  
 Matches 240; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 SGGCFWNGHLYREDQTSAPGLRCLNWLDAQSLASAPVSGAGNHSYCRNPDEDPGPW 60  
 DB 22 SGGCFWNGHLYREDQTSAPGLRCLNWLDAQSLASAPVSGAGNHSYCRNPDEDPGPW 81

QY 61 CYYSGEAGVPEKPCEDLRCPETTSQALPAFTTEIQEASEGPGADVQVAFANALPARS 120  
 DB 82 CYYSGEAGVPEKPCEDLRCPETTSQALPAFTTEIQEASEGPGADVQVAFANALPARS 141

QY 121 EAAAVQPVIGISORVRMNSKEKDLGTGLGVLTGVTWVHIIAIGAGIILGYSYKRGKDLK 180  
 DB 142 EAAAVQPVIGISORVRMNSKEKDLGTGLGVLTGVTWVHIIAIGAGIILGYSYKRGKDLK 201

QY 181 EQHDQKVCEREMQRIITPLSAFTNPTCEIVDEKTVVHTSQTPVDPQEGSTPLMQAGTP 240  
 DB 202 EQHDQKVCEREMQRIITPLSAFTNPTCEIVDEKTVVHTSQTPVDPQEGSTPLMQAGTP 261

QY 241 GA 242  
 DB 262 GA 263

RESULT 8  
 AAY05220  
 ID AAY05220 standard; protein; 286 AA.  
 XX  
 AC AAY05220;  
 XX  
 DT 17-JUN-1999 (first entry)  
 DE Kringle1 protein sequence.  
 XX  
 KW Kringle1; cancer; inflammation; autoimmunity; rheumatoid arthritis; AIDS;  
 KW CNS inflammation; cerebellar degeneration; Alzheimer's disease; asthma;  
 KW Parkinson's disease; multiple sclerosis; amyotrophic lateral sclerosis;  
 KW neurological abnormality; ischaemia reperfusion injury; ischaemic injury;  
 KW cardiovascular disease; kidney disease; liver disease; aplastic anaemia;  
 KW myocardial infarction; hypertension; hypertension; allergy; infection;  
 KW myelodysplastic syndrome; haematologic abnormality; diagnosis; therapy;  
 KW male pattern baldness.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO9911788-A1.  
 XX  
 PD 11-MAR-1999.  
 XX  
 PF 02-SEP-1998; 98WO-US018270.  
 XX  
 PR 02-SEP-1997; 97US-0056032P.  
 PR 01-SEP-1998; 98US-00144889.  
 XX  
 PA (SMK ) SMITHKLINE BEECHAM CORP.  
 XX  
 PI Albione EF, Kikly KK;  
 XX  
 DR WPI; 1999-214707/18.  
 DR N-PSDB; AAX28355.  
 XX  
 PT New kringle1 polypeptides and polynucleotides.  
 XX  
 PS Claim 14; Page 33; 42pp; English.  
 XX

This sequence is a Kringle1 polypeptide of the invention. The kringle1 polypeptides (I) are used to screen for agonists and antagonists.  
 CC Agonists are used to treat subjects in need of enhanced activity or



expression of (I). Antagonists are used to treat subjects having need to inhibit the activity or expression of (I). The methods can be used to treat conditions such as cancer, inflammation, autoimmunity, allergy, asthma, rheumatoid arthritis, CNS inflammation, cerebellar degeneration, Alzheimer's disease, Parkinson's disease, multiple sclerosis, amyotrophic lateral sclerosis, head injury damage and other neurological abnormalities, ischaemia reperfusion injury, cardiovascular disease, kidney disease, liver disease, ischaemic injury, myocardial infarction, hypertension, hypotension, AIDS, myelodysplastic syndromes and other haematologic abnormalities, aplastic anaemia, male pattern baldness, and bacterial, fungal, protozoan and viral infections. The kringiel polypeptides may also be used to generate antibodies. Determining the presence or absence of mutations in, and analysing for the presence or absence of expression of, kringiel polynucleotides can be used to diagnose a disease or susceptibility to a disease related to expression or activity of kringiel proteins. The polynucleotides may also be used for chromosome identification, and mapping

XX SQ Sequence 286 AA;

Query Match 82.5%; Score 1063.5; DB 2; Length 286;  
Best Local Similarity 86.2%; Pred. No. 3.8e-92;  
Matches 206; Conservative 5; Mismatches 27; Indels 1; Gaps 1;  
QY 1 SGGCFWNGHLYREDQTSPPAGLRCLNWLDAQSLASAPVSGAGNHSYCRNPDEDPGWP 60  
Db 22 SGGCFWNGHLYREDQTSPPAGLRCLNWLDAQSGPASAPVSGADNHSYCRNPDEDPGWP 81  
QY 61 CVVSGEAGVPEKPCEDLRCPETTSQALPAP-TTEIQEASGEGGADDEVQVAPANALPAR 119  
Db 82 CVVSGEAGVPEKPCEDLRCPETTSQALPAFHDSNRKLEGGGADDEVQVAPANALPAR 141  
QY 120 SEAAVQPVGIGISQVRMNSKXKDLGTGLVGLITMVIITAGIILGYSYKRGKDL 179  
Db 142 SEAAVQPVGIGISQVRMNSKXKDLGTGLVGLITMVIITAGIILGYSYKRGKDL 201  
QY 180 KEQHQVQVCEQVQVITPLSAFTNPTCEIVDEKTVVHTSQTPVDPQSGPLMGQAG 238  
Db 202 KEQHQVQVCEQVQVITPLSAFTNPTCEIVDEKTVVHTSQTPVDPQSGPLMGQAG 260

RESULT 9

ABR40487  
ID ABR40487 standard; protein; 146 AA.

XX ABR40487;

XX 13-JUN-2003 (first entry)

XX Human secreted protein #SEQ ID 237.

Human; secreted protein; anti-HIV; nootropic; neuroprotective;  
antianginal; immunosuppressive; immunomodulator; cytostatic; cardiant;  
hepatotropic; antiinflammatory; antiallergic; antidiabetic; cardiant;  
gastrointestinal; antinfertility; nephrotropic; virucide; hypotensive;  
vasotrophic; dermatological; osteopathic; antiarthritic; antiparkinsonian;  
antiaesthetic; antipsoriatic; cerebroprotective; antibacterial;  
fungicide; antiparasitic; protein therapy; gene therapy; cancer;  
hyperproliferative disorder; leukaemia; autoimmune disorder;  
immunodeficiency; HIV; infertility; uterine disorders; endometriosis;  
amenorrhea; ocular disorder; neurological disorder; wound healing;  
Huntington's disease; gastrointestinal disorder; inflammatory disease;  
Crohn's disease; kidney disorder; renal failure; diabetes; obesity;  
cirrhosis; hepatitis; cardiovascular disorder; angina; osteoarthritis;  
rheumatoid arthritis; psoriasis; multiple sclerosis; asthma;  
respiratory disorder; infectious disease; chromosome identification;  
food additive; nutrition.

XX Homo sapiens.

XX WO200268528-A1.

XX 06-SEP-2002.

XX 21-FEB-2002; 2002WO-US005301.

XX 23-FEB-2001; 2001US-0270625P.

PR 12-JUN-2001; 2001US-0304417P.

XX (HUMA-) HUMAN GENOME SCI INC.

PI Rosen CA, Komatsoulis G, Baker KP, Fiscella M, Moore PA, Wei P;

PI Duan DR, Shi Y, Gupta R;

XX WPI; 2002-750417/81.

DR N-PSDB; ABZ82342.

XX New human secreted proteins and nucleic acids, useful for preventing,

PT treating, or ameliorating a medical condition, e.g. cancer, autoimmune

PT disorder, HIV, Parkinson's disease, Alzheimer's disease, diabetes,

PT obesity or cirrhosis.

XX Claim 11; Page 800; 873pp; English.

XX The invention relates to novel human secreted proteins and the genes

CC encoding them. Genes and proteins of the invention may be useful for

CC preventing, treating or ameliorating medical conditions e.g. by protein

CC or gene therapy. These conditions include cancer and hyperproliferative

CC disorders, immune cell proliferative disorders (e.g. leukaemia),

CC autoimmune disorders, immunodeficiencies (e.g. HIV infection),

CC infertility, placental and uterine disorders (e.g. endometriosis),

CC amenorrhea, ocular disorders, neurological disorders (e.g. Huntington's

CC disease), wound healing, gastrointestinal system disorders, particularly

CC inflammatory diseases (e.g. Crohn's disease), kidney disorders (e.g.

CC renal failure), diabetes, obesity, cirrhosis, hepatitis, cardiovascular

CC disorders (e.g. anginal), rheumatoid arthritis, osteoarthritis, psoriasis,

CC multiple sclerosis, respiratory disorders (e.g. asthma), or infectious

CC diseases caused by bacterial, parasitic, viral or fungal agents. The

CC nucleic acids are also useful for chromosome identification, radiation

CC hybrid mapping or long-range restriction mapping. The polypeptide,

CC polynucleotide, agonist or antagonist may also be used as a food additive

CC or preservative to increase or decrease storage capabilities, fat content

CC or other nutritional components. The sequences given in records ABR40409-

CC ABR40590 and ABZ82464-ABZ82611 represent human secreted proteins and the

CC genes encoding them

XX Sequence 146 AA;

XX Query Match 50.3%; Score 648; DB 5; Length 146;

XX Best Local Similarity 93.6%; Pred. No. 3.5e-53;

XX Matches 117; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

QY 1 SGGCFWNGHLYREDQTSPPAGLRCLNWLDAQSLASAPVSGAGNHSYCRNPDEDPGWP 60

Db 22 SGGCFWNGHLYREDQTSPPAGLRCLNWLDAQSLASAPVSGAGNHSYCRNPDEDPGWP 81

QY 61 CVVSGEAGVPEKPCEDLRCPETTSQALPAPTTTIEQASGEGGADDEVQVAPANALPAR 120

Db 82 CVVSGEAGVPEKPCEDLRCPETTSQALPAPTTTIEQASGEGGADDEVQVAPATXPPLGS 141

QY 121 EAAAV 125

Db 142 XAXAV 146

XX RESULT 10

XX ABR40561

XX ID ABR40561 standard; protein; 146 AA.

XX ABR40561;

XX 13-JUN-2003 (first entry)

XX Human secreted protein #SEQ ID 311.

XX Human; secreted protein; anti-HIV; nootropic; neuroprotective;

KW anti-anginal; immunosuppressive; immunomodulator; cytotactic; cardiant;  
KW hepatotropic; anti-inflammatory; antiallergic; antidiabetic; hypotensive;  
KW gastrointestinal; antifertility; nephrotropic; virucide; antiparkinsonian;  
KW vasotropic; dermatological; osteopathic; antiarthritic; antibacterial;  
KW antiasthmatic; antiparasitic; cerbroprotective; cancer;  
KW fungicide; antiparasitic; protein therapy; gene therapy; cancer;  
KW hyperproliferative disorder; leukemia; autoimmune disorder;  
KW immunodeficiency; HIV; infertility; uterine disorders; endometriosis;  
KW amenorrhea; ocular disorder; neurological disorder; wound healing;  
KW Huntington's disease; gastrointestinal disorder; inflammatory disease;  
KW Crohn's disease; kidney disorder; renal failure; diabetes; obesity;  
KW cirrhosis; hepatitis; cardiovascular disorder; angina; osteoarthritis;  
KW rheumatoid arthritis; psoriasis; multiple sclerosis; asthma;  
KW respiratory disorder; infectious disease; chromosome identification;  
KW food additive; nutrition.  
XX  
OS Homo sapiens.  
XX  
FN WO200268628-A1.  
XX  
PD 06-SEP-2002.  
XX  
PF 21-FEB-2002; 2002WO-US005301.  
XX  
PR 23-FEB-2001; 2001US-0270625P.  
XX  
PR 12-JUL-2001; 2001US-0304417P.  
XX  
PA (HUMA-) HUMAN GENOME SCI INC.  
XX  
XX Rosen CA, Komatsoulis G, Baker KP, Fiscella M, Moore PA, Wei P;  
PI Duan DR, Shi Y, Gupta R;  
XX  
XX WI; 2002-750417/81.  
XX  
XX New human secreted proteins and nucleic acids, useful for preventing,  
PT treating, or ameliorating a medical condition, e.g. cancer, autoimmune  
PT disorder, HIV, Parkinson's disease, Alzheimer's disease, diabetes,  
PT obesity or cirrhosis.  
XX  
PS Claim 11; Page 850; 873pp; English.  
XX  
CC The invention relates to novel human secreted proteins and the genes  
CC encoding them. Genes and proteins of the invention may be useful for  
CC preventing, treating or ameliorating medical conditions e.g. by protein  
CC or gene therapy. These conditions include cancer and hyperproliferative  
CC disorders, immune cell proliferative disorders (e.g. leukemia),  
CC autoimmune disorders, immunodeficiencies (e.g. HIV infection),  
CC infertility, placental and uterine disorders (e.g. endometriosis),  
CC amenorrhea, ocular disorders, neurological disorders (e.g. Huntington's  
CC disease), wound healing, gastrointestinal system disorders, particularly  
CC inflammatory diseases (e.g. Crohn's disease), kidney disorders (e.g.  
CC renal failure), diabetes, obesity, cirrhosis, hepatitis, cardiovascular  
CC disorders (e.g. angina), rheumatoid arthritis, osteoarthritis, psoriasis,  
CC multiple sclerosis, respiratory disorders (e.g. asthma), or infectious  
CC diseases caused by bacterial, parasitic, viral or fungal agents. The  
CC nucleic acids are also useful for chromosome identification, radiation  
CC hybrid mapping or long-range restriction mapping. The polypeptide,  
CC polynucleotide, agonist or antagonist may also be used as a food additive  
CC or preservative to increase or decrease storage capabilities, fat content  
CC or other nutritional components. The sequences given in records ABR40409-  
CC ABR40590 and AB282464-AB282611 represent human secreted proteins and the  
CC genes encoding them  
XX  
SQ Sequence 146 AA;

Query Match 50.3%; Score 648; DB 5; Length 146;  
Best Local Similarity 93.6%; Pred. No. 3.5e-53;  
Matches 117; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

Qy 1 SGGCFWNGHLYREDQTSFAPGLCLNWLDAQSLASAPVSGAGNHSYCRNPDDPRGW 60  
Db 22 SGGCFWNGHLYREDQTSFAPGLCLNWLDAQSLASAPVSGAGNHSYCRNPDDPRGW 81

Qy 61 CYVSGEAGVPEKPCEDLCPCETTSQALPAFTTEIOEASEGPGADEVQVFPANALPARS 120  
Db 82 CYVSGEAGVPEKPCEDLCPCETTSQALPAFTTEIOEASEGPGADEVQVFXPATPFLGS 141  
Qy 121 EAAAV 125  
Db 142 XAXAV 146  
RESULT 11  
ABR42624  
ID ABR42624 standard; protein; 81 AA.  
XX AC ABR42624;  
XX  
DT 26-AUG-2003 (first entry)  
XX  
XX Human kringle containing protein.  
XX  
KW Human; abrogen; kringle; angiogenesis; inhibitor; tumour; metastasis;  
KW cytotactic; gene therapy; expressed sequence tag; EST.  
XX  
XX Homo sapiens.  
OS  
XX  
FH Key Location/Qualifiers  
FT Misc-difference 4 /note= "the identity of this residue is unclear in Fig 2"  
FT Misc-difference 6 /note= "the identity of this residue is unclear in Fig 2"  
FT Misc-difference 7 /note= "the identity of this residue is unclear in Fig 2"  
FT Misc-difference 16 /note= "the identity of this residue is unclear in Fig 2"  
FT Misc-difference 29 /note= "the identity of this residue is unclear in Fig 2"  
FT Misc-difference 32 /note= "the identity of this residue is unclear in Fig 2"  
FT Misc-difference 46 /note= "the identity of this residue is unclear in Fig 2"  
FT Domain 50..56 /note= "kringle domain"  
FT Misc-difference 57 /note= "the identity of this residue is unclear in Fig 2"  
FT Misc-difference 62 /note= "the identity of this residue is unclear in Fig 2"  
FT Misc-difference 70 /note= "the identity of this residue is unclear in Fig 2"  
FT Misc-difference 72 /note= "the identity of this residue is unclear in Fig 2"  
FT Misc-difference 77 /note= "the identity of this residue is unclear in Fig 2"  
XX WO2003042354-A2.  
XX  
PD 22-MAY-2003.  
XX  
XX 04-SEP-2002; 2002WO-US027885.  
XX  
XX 04-SEP-2001; 2001US-0316300P.  
XX (AVET ) AVENTIS PHARM INC.  
XX  
XX Nesbit M, Fong TC, Brockstedt D;  
XX  
XX WPI; 2003-449566/42.  
XX  
XX New abrogen polypeptide, useful for treating an angiogenesis related  
PT diseases e.g. tumor metastasis.  
XX  
XX Disclosure; Fig 2; 95pp; English.  
XX  
XX The present sequence is the protein sequence of a hypothetical kringle-  
CC containing protein encoded by an expressed sequence tag. The invention

CC relates to novel abrogen polypeptides that are derived from kringle-  
CC containing proteins. The abrogens are potent inhibitors of endothelial  
CC proliferation and angiogenesis. They are capable of inhibiting or  
CC reducing cell proliferation induced by both basic fibroblast growth  
CC factor and vascular endothelial growth factor in a specific endothelial  
CC cell proliferation assay. Vectors that expressed abrogen polypeptides in  
CC vivo were shown to reduce tumour metastasis in 2 lung cancer models. The  
CC invention provides abrogen polypeptides and polynucleotides, and methods  
CC of using these to treat an angiogenesis-related disease or disorder, e.g.  
CC tumour metastasis (claimed)  
XX Sequence 81 AA;  
SQ

Query Match 29.9%; Score 385; DB 7; Length 81;  
Best Local Similarity 82.7%; Pred. No. 1.3e-28;  
Matches 67; Conservative 0; Mismatches 14; Indels 0; Gaps 0;

QY 2 GGCWNGHLYREDOTSPAPGLRCNLWLDAQGLASAPVSGAGNHSYCRNPDEDPGFWC 61  
DB 1 GGCWXXGHLHYREDQXSPAFGLRCNLWLXQXGLASAPVSGAGNHSYCRNPDEDPGFWC 60

QY 62 YVSGEAGVPEKRPCEDLRCE 82  
DB 61 YXSGEAGVPEKPCEDXRCE 81

RESULT 12  
AAM18800  
ID AAM18800 standard; protein; 66 AA.  
XX  
AC AAM18800;

DT 12-OCT-2001 (first entry)  
XX  
DE Peptide #5234 encoded by probe for measuring cervical gene expression.

XX Probe; human; microarray; gene expression; cervical epithelial cell;  
KW cervical cancer.  
XX Homo sapiens.

XX WO200157278-A2.  
XX  
XX 09-AUG-2001.

XX 30-JAN-2001; 2001WO-US000670.  
XX  
XX 04-FEB-2000; 2000US-0180312P.

XX 26-MAY-2000; 2000US-0207456P.  
XX 30-JUN-2000; 2000US-00608408.  
XX 03-AUG-2000; 2000US-00632366.

XX 21-SEP-2000; 2000US-0234687P.  
XX 27-SEP-2000; 2000US-0236359P.  
XX 04-OCT-2000; 2000GB-00024263.

XX (MOLE-) MOLECULAR DYNAMICS INC.

XX Penn SG, Hanzel DK, Chen W, Rank DR;  
XX WPI; 2001-488901/53.

XX Human genome-derived single exon nucleic acid probes useful for analyzing  
XX gene expression in human cervical epithelial cells.

XX Claim 27; SEQ ID NO 23626; 487pp; English.

XX The present invention relates to human single exon nucleic acid probes  
XX (SENPs; see AAI10068-AAI28459). The present sequence is a peptide encoded  
XX by one such probe. The SENPs are derived from human HeLa cells. The SENPs  
XX can be used to produce a single exon microarray, which can be used for  
XX measuring human gene expression in a sample derived from human cervical  
XX epithelial cells. By measuring gene expression, the probes are therefore  
XX useful in grading and/or staging of diseases of the cervix, notably

CC cervical cancer. Note: The sequence data for this patent did not form  
CC part of the printed specification, but was obtained in electronic format  
CC directly from WIPO at ftp.wipo.int/pub/published\_pct\_sequences  
XX  
XX Sequence 66 AA;

Query Match 25.0%; Score 322; DB 4; Length 66;  
Best Local Similarity 100.0%; Pred. No. 9e-23;  
Matches 66; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 83 TTSQLPAFTTEIQEASGEGGDEVQVFAPANALPARSEAAAVQPVIGISQVRMNSKEK 142  
DB 1 TTSQLPAFTTEIQEASGEGGDEVQVFAPANALPARSEAAAVQPVIGISQVRMNSKEK 60

QY 143 KDLGTL 148  
DB 61 KDLGTL 66

RESULT 13

ABB37905  
ID ABB37905 standard; peptide; 66 AA.

XX  
AC ABB37905;

XX 04-FEB-2002 (first entry)

XX Peptide #5411 encoded by human foetal liver single exon probe.

XX Human; foetal liver; gene expression; single exon nucleic acid probe.

XX Homo sapiens.

XX WO200157277-A2.

XX 09-AUG-2001.

XX 30-JAN-2001; 2001WO-US000669.

XX 04-FEB-2000; 2000US-0180312P.

XX 26-MAY-2000; 2000US-0207456P.

XX 30-JUN-2000; 2000US-00608408.

XX 03-AUG-2000; 2000US-00632366.

XX 21-SEP-2000; 2000US-0234687P.

XX 27-SEP-2000; 2000US-0236359P.

XX 04-OCT-2000; 2000GB-00024263.

XX (MOLE-) MOLECULAR DYNAMICS INC.

XX Penn SG, Hanzel DK, Chen W, Rank DR;

XX WPI; 2001-483447/52.

XX Human genome-derived single exon nucleic acid probes useful for analyzing  
XX gene expression in human foetal liver.

XX Claim 27; SEQ ID NO 30540; 639pp + Sequence Listing; English.

XX The invention relates to a single exon nucleic acid probe for measuring  
XX human gene expression in a sample derived from human foetal liver. The  
XX single exon nucleic acid probes may be used for predicting, measuring and  
XX displaying gene expression in samples derived from human foetal liver. The  
XX present sequence is a peptide encoded by a single exon nucleic acid probe  
XX of the invention. Note: The sequence data for this patent did not form  
XX part of the printed specification, but was obtained in electronic format  
XX directly from WIPO at ftp.wipo.int/pub/published\_pct\_sequences  
XX Sequence 66 AA;

Query Match 25.0%; Score 322; DB 4; Length 66;  
Best Local Similarity 100.0%; Pred. No. 9e-23;  
Matches 66; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 83 TTSQLPAPFTTEIQASEGPGADEVQVPAPANALPARSEAAVQPVIGISORVRMNSKEK 142  
Db 1 TTSQLPAPFTTEIQASEGPGADEVQVPAPANALPARSEAAVQPVIGISORVRMNSKEK 60  
QY 143 KDLGTL 148  
Db 61 KDLGTL 66  
RESULT 14  
AAM31314  
ID AAM31314 standard; protein; 66 AA.  
XX AC  
XX AAM31314;  
XX 17-OCT-2001. (first entry)  
XX Peptide #5351 encoded by probe for measuring placental gene expression.  
DE XX  
XX Probe; microarray; human; placenta; antenatal diagnosis;  
KW genetic disorder.  
KW Homo sapiens.  
XX OS  
XX WO200157272-A2.  
PN 27-SEP-2000; 2000US-0234687P.  
XX 09-AUG-2001.  
PD 30-JAN-2001; 2001WO-US000663.  
XX 04-FEB-2000; 2000US-0180312P.  
XX 26-MAY-2000; 2000US-0207456P.  
PR 30-JUN-2000; 2000US-00608408.  
PR 03-AUG-2000; 2000US-00632366.  
PR 21-SEP-2000; 2000US-0234687P.  
PR 27-SEP-2000; 2000US-0236359P.  
PR 04-OCT-2000; 2000GB-00024263.  
XX (MOLE-) MOLECULAR DYNAMICS INC.  
PA Penn SG, Hanzel DK, Chen W, Rank DR;  
XX WPI; 2001-488897/53.  
XX Human genome-derived single exon nucleic acid probes useful for analyzing  
DR gene expression in human placenta.  
XX Claim 27; SEQ ID NO 31583; 654pp; English.  
XX The present invention relates to single exon nucleic acid probes (SENP:  
CC see AAI31315-AAI57546). The present sequence is a peptide encoded by one  
CC such probe. The probes are useful for producing a microarray for  
CC predicting, measuring and displaying gene expression in samples derived  
CC from human placenta. The probes are useful for antenatal diagnosis of  
CC human genetic disorders  
XX Sequence 66 AA;  
SQ Query Match 25.0%; Score 322; DB 4; Length 66;  
Best Local Similarity 100.0%; Pred. No. 9e-23;  
Matches 66; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 83 TTSQLPAPFTTEIQASEGPGADEVQVPAPANALPARSEAAVQPVIGISORVRMNSKEK 142  
Db 1 TTSQLPAPFTTEIQASEGPGADEVQVPAPANALPARSEAAVQPVIGISORVRMNSKEK 60  
QY 143 KDLGTL 148  
Db 61 KDLGTL 66  
RESULT 15  
ABB23159

ID XX ABB23159 standard; protein; 66 AA.  
AC ABB23159;  
XX 23-JAN-2002 (first entry)  
DT Protein #5158 encoded by probe for measuring heart cell gene expression.  
XX Human; gene expression; heart; microarray; vascular system;  
DE cardiovascular disease; hypertension; cardiac arrhythmia;  
KW congenital heart disease.  
KW Homo sapiens.  
XX OS  
XX WO200157274-A2.  
PN 09-AUG-2001.  
PD 30-JAN-2001; 2001WO-US000666.  
XX 04-FEB-2000; 2000US-0180312P.  
XX 26-MAY-2000; 2000US-0207456P.  
PR 30-JUN-2000; 2000US-00608408.  
PR 03-AUG-2000; 2000US-00632366.  
PR 21-SEP-2000; 2000US-0234687P.  
PR 27-SEP-2000; 2000US-0236359P.  
PR 04-OCT-2000; 2000GB-00024263.  
XX (MOLE-) MOLECULAR DYNAMICS INC.  
PA Penn SG, Hanzel DK, Chen W, Rank DR;  
XX WPI; 2001-488899/53.  
XX Single exon nucleic acid probes for analyzing gene expression in human  
PT hearts.  
XX Claim 15; SEQ ID NO 24929; 530pp; English.  
XX The present invention relates to single exon nucleic acid probes for  
CC measuring human gene expression in a sample derived from human heart (see  
CC ABA21535-ABA41305). The present sequence is a protein encoded by one such  
CC probe. The probes may be used for predicting, measuring and displaying  
CC gene expression in samples derived from the human heart via microarrays.  
CC By measuring gene expression, the probes are useful for predicting of the  
CC diagnosing, grading, staging, monitoring and prognosing diseases of the  
CC human heart and vascular system e.g. cardiovascular disease,  
CC hypertension, cardiac arrhythmias and congenital heart disease. Note: The  
CC sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pat\_sequences  
XX Sequence 66 AA;  
SQ Query Match 25.0%; Score 322; DB 4; Length 66;  
Best Local Similarity 100.0%; Pred. No. 9e-23;  
Matches 66; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 83 TTSQLPAPFTTEIQASEGPGADEVQVPAPANALPARSEAAVQPVIGISORVRMNSKEK 142  
Db 1 TTSQLPAPFTTEIQASEGPGADEVQVPAPANALPARSEAAVQPVIGISORVRMNSKEK 60  
QY 143 KDLGTL 148  
Db 61 KDLGTL 66  
Search completed: March 17, 2004, 07:03:42  
Job time : 61.8843 secs

Result No.	Score	Query		Length	DB	ID	Description
		Match	%				
1	1289	100.0-	263	9	US-09-084-491A-2		Sequence 2, Appli
2	1289	100.0-	263	13	US-10-102-704-2		Sequence 2, Appli
3	1289	100.0-	263	13	US-10-057-951-2		Sequence 2, Appli
4	1286	99.8	263	14	US-10-210-951-44		Sequence 44, Appli
5	1286	99.8	263	14	US-10-211-884-44		Sequence 44, Appli
6	332	25.0	66	9	US-09-864-751-38457		Sequence 38457, A
7	154.5	12.0	527	10	US-09-987-457-18		Sequence 18, Appl
8	154.5	12.0	527	10	US-09-987-455-19		Sequence 19, Appl
9	154.5	12.0	527	15	US-10-360-101-203		Sequence 203, App
10	154.5	12.0	562	9	US-09-969-271-7		Sequence 7, Appli
11	154.5	12.0	562	9	US-09-974-299-145		Sequence 145, App
12	154.5	12.0	562	12	US-10-411-037-26		Sequence 26, Appli
13	154.5	12.0	562	14	US-10-193-656-8		Sequence 8, Appli
14	154.5	12.0	562	15	US-10-443-701-4		Sequence 4, Appli
15	154.5	12.0	650	15	US-10-401-077-1		Sequence 1, Appli

```
/ MOLECULE TYPE: protein
US-09-084-491A-2

Query Match      100.0%; Score 1289; DB 9; Length 263;
Best Local Similarity 100.0%; Pred. No. 9e-116;
Matches 242; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SGGCFWNGHLYREDQTSAPGLRCLNWLDAQSGLASAPVSGAGNHSYCRNPDEDPRGPW 60
DB 22 SGGCFWNGHLYREDQTSAPGLRCLNWLDAQSGLASAPVSGAGNHSYCRNPDEDPRGPW 81

QY 61 CYSGEAGVPEKPCEDLRCPTTSQALPAFTTEIQEASEGPGADEVQVFAPANALPARS 120
DB 82 CYSGEAGVPEKPCEDLRCPTTSQALPAFTTEIQEASEGPGADEVQVFAPANALPARS 141

QY 121 EAAAVQPVIGISQVRVNSKEKDLGTLGYVLGITMWVIIIAIGAGIILGYSYKRGKDLK 180
DB 142 EAAAVQPVIGISQVRVNSKEKDLGTLGYVLGITMWVIIIAIGAGIILGYSYKRGKDLK 201

QY 181 EHQDKVCEREMQRIITPLSAFTNPTCEIVDEKTVVHTSQTVPDPQEGSTPLMQAGTP 240
DB 202 EHQDKVCEREMQRIITPLSAFTNPTCEIVDEKTVVHTSQTVPDPQEGSTPLMQAGTP 261

QY 241 GA 242
DB 262 GA 263

RESULT 2
US-10-102-704-2
; Sequence 2, Application US/10102704
; Publication No. US20020164768A1
; GENERAL INFORMATION:
; APPLICANT: Moore et al.
; TITLE OF INVENTION: Tissue Plasminogen Activator-Like Protein
; FILE REFERENCE: P378C1
; CURRENT APPLICATION NUMBER: US/10/102,704
; CURRENT FILING DATE: 2002-03-22
; PRIOR APPLICATION NUMBER: 09/084,491
; PRIOR FILING DATE: 1998-05-27
; PRIOR APPLICATION NUMBER: 60/048,000
; PRIOR FILING DATE: 1997-05-28
; NUMBER OF SEQ ID NOS: 15
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 2
; LENGTH: 263
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-102-704-2

Query Match      100.0%; Score 1289; DB 13; Length 263;
Best Local Similarity 100.0%; Pred. No. 9e-116;
Matches 242; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SGGCFWNGHLYREDQTSAPGLRCLNWLDAQSGLASAPVSGAGNHSYCRNPDEDPRGPW 60
DB 22 SGGCFWNGHLYREDQTSAPGLRCLNWLDAQSGLASAPVSGAGNHSYCRNPDEDPRGPW 81

QY 61 CYSGEAGVPEKPCEDLRCPTTSQALPAFTTEIQEASEGPGADEVQVFAPANALPARS 120
DB 82 CYSGEAGVPEKPCEDLRCPTTSQALPAFTTEIQEASEGPGADEVQVFAPANALPARS 141

QY 121 EAAAVQPVIGISQVRVNSKEKDLGTLGYVLGITMWVIIIAIGAGIILGYSYKRGKDLK 180
DB 142 EAAAVQPVIGISQVRVNSKEKDLGTLGYVLGITMWVIIIAIGAGIILGYSYKRGKDLK 201

QY 181 EHQDKVCEREMQRIITPLSAFTNPTCEIVDEKTVVHTSQTVPDPQEGSTPLMQAGTP 240
DB 202 EHQDKVCEREMQRIITPLSAFTNPTCEIVDEKTVVHTSQTVPDPQEGSTPLMQAGTP 261

QY 241 GA 242
DB 262 GA 263

RESULT 4
US-10-210-951-44
; Sequence 44, Application US/10210951
; Publication No. US20030170228A1
; GENERAL INFORMATION:
; APPLICANT: Ashkenazi, Avi J.
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Hillan, Kenneth J.
; APPLICANT: Marsters, Scott A.
; APPLICANT: Pan, James
; APPLICANT: Pitti, Robert M.
; APPLICANT: Roy, Margaret Ann
; APPLICANT: Smith, Victoria
; APPLICANT: Stone, Donna M.
; APPLICANT: Watanabe, Colin K.
; APPLICANT: Wood, William I.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE TREATMENT OF TUMOR
; FILE REFERENCE: P2931R1C1
; CURRENT APPLICATION NUMBER: US/10/210,951
; CURRENT FILING DATE: 2002-08-02
```

; PRIOR APPLICATION NUMBER: 60/014699  
; CURRENT FILING DATE: 1996-04-01  
; PRIOR APPLICATION NUMBER: 60/026943  
; PRIOR FILING DATE: 1996-09-23  
; PRIOR APPLICATION NUMBER: 60/059121  
; PRIOR FILING DATE: 1997-07-17  
; PRIOR APPLICATION NUMBER: 60/059352  
; PRIOR FILING DATE: 1997-09-19  
; PRIOR APPLICATION NUMBER: 60/062037  
; PRIOR FILING DATE: 1997-10-10  
; PRIOR APPLICATION NUMBER: 60/063755  
; PRIOR FILING DATE: 1997-10-17  
; PRIOR APPLICATION NUMBER: 60/063045  
; PRIOR FILING DATE: 1997-10-24  
; PRIOR APPLICATION NUMBER: 60/063046  
; PRIOR FILING DATE: 1997-10-24  
; PRIOR APPLICATION NUMBER: 60/066511  
; PRIOR FILING DATE: 1997-11-24  
; PRIOR APPLICATION NUMBER: 60/066772  
; PRIOR FILING DATE: 1997-11-24  
; Remaining Prior Application data removed - See File Wrapper or PALM.  
; NUMBER OF SEQ ID NOS: 258  
; SEQ ID NO 44  
; LENGTH: 263  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-210-951-44

Query Match 99.8%; Score 1286; DB 14; Length 263;  
Best Local Similarity 99.6%; Pred. No. 1.7e-115;  
Matches 241; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 SGGCFWNGHLYREDQTSPPAGLRCINLWLDQAOSGLASAPVSGAGNHSCRNPDPRGPW 60  
DB 22 SGGCFWNGHLYREDQTSPPAGLRCINLWLDQAOSGLASAPVSGAGNHSCRNPDPRGPW 81  
QY 61 CYYSGEAGVPEKPCEDLRCPETTSQALPAFTTEIOEASEGPGADEVQVFAPANALPARS 120  
DB 82 CYYSGEAGVPEKPCEDLRCPETTSQALPAFTTEIOEASEGPGADEVQVFAPANALPARS 141  
QY 121 EAAAVQPVIGISQVRVNSKEKDLGTLGVLGITMMVIIIAAGIILGYSKGKDLK 180  
DB 142 EAAAVQPVIGISQVRVNSKEKDLGTLGVLGITMMVIIIAAGIILGYSKGKDLK 201  
QY 181 EQHDQKVCEREMQRIITLPSAFTNPTCEIVDEKTVVHTSQTVPDPOEGSTPLMGQAGTP 240  
DB 202 EQHDQKVCEREMQRIITLPSAFTNPTCEIVDEKTVVHTSQTVPDPOEGCTTPLMGQAGTP 261  
QY 241 GA 242  
DB 262 GA 263

RESULT 5  
US-10-211-884-44  
; Sequence 44, Application US/10211884  
; Publication No. US20030175900A1  
; GENERAL INFORMATION:  
; APPLICANT: Ashkenazi, Avi J.  
; APPLICANT: Goddard, Audrey  
; APPLICANT: Godowski, Paul J.  
; APPLICANT: Gurney, Austin L.  
; APPLICANT: Hillan, Kenneth J.  
; APPLICANT: Marsters, Scot A.  
; APPLICANT: Pan, James  
; APPLICANT: Pitti, Robert M.  
; APPLICANT: Roy, Margaret Ann  
; APPLICANT: Smith, Victoria  
; APPLICANT: Stone, Donna M.  
; APPLICANT: Watanabe, Colin K.  
; APPLICANT: Wood, William I.  
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE TREATMENT OF TUMOR  
; FILE REFERENCE: P2931R1C1

; CURRENT APPLICATION NUMBER: US/10/211,884  
; CURRENT FILING DATE: 2002-08-02  
; PRIOR APPLICATION NUMBER: 60/014699  
; PRIOR FILING DATE: 1996-04-01  
; PRIOR APPLICATION NUMBER: 60/026943  
; PRIOR FILING DATE: 1996-09-23  
; PRIOR APPLICATION NUMBER: 60/059121  
; PRIOR FILING DATE: 1997-07-17  
; PRIOR APPLICATION NUMBER: 60/059352  
; PRIOR FILING DATE: 1997-09-19  
; PRIOR APPLICATION NUMBER: 60/062037  
; PRIOR FILING DATE: 1997-10-10  
; PRIOR APPLICATION NUMBER: 60/063755  
; PRIOR FILING DATE: 1997-10-17  
; PRIOR APPLICATION NUMBER: 60/063045  
; PRIOR FILING DATE: 1997-10-24  
; PRIOR APPLICATION NUMBER: 60/063046  
; PRIOR FILING DATE: 1997-10-24  
; PRIOR APPLICATION NUMBER: 60/066511  
; PRIOR FILING DATE: 1997-11-24  
; PRIOR APPLICATION NUMBER: 60/066772  
; PRIOR FILING DATE: 1997-11-24  
; Remaining Prior Application data removed - See File Wrapper or PALM.  
; NUMBER OF SEQ ID NOS: 258  
; SEQ ID NO 44  
; LENGTH: 263  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-211-884-44

Query Match 99.8%; Score 1286; DB 14; Length 263;  
Best Local Similarity 99.6%; Pred. No. 1.7e-115;  
Matches 241; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 SGGCFWNGHLYREDQTSPPAGLRCINLWLDQAOSGLASAPVSGAGNHSCRNPDPRGPW 60  
DB 22 SGGCFWNGHLYREDQTSPPAGLRCINLWLDQAOSGLASAPVSGAGNHSCRNPDPRGPW 81  
QY 61 CYYSGEAGVPEKPCEDLRCPETTSQALPAFTTEIOEASEGPGADEVQVFAPANALPARS 120  
DB 82 CYYSGEAGVPEKPCEDLRCPETTSQALPAFTTEIOEASEGPGADEVQVFAPANALPARS 141  
QY 121 EAAAVQPVIGISQVRVNSKEKDLGTLGVLGITMMVIIIAAGIILGYSKGKDLK 180  
DB 142 EAAAVQPVIGISQVRVNSKEKDLGTLGVLGITMMVIIIAAGIILGYSKGKDLK 201  
QY 181 EQHDQKVCEREMQRIITLPSAFTNPTCEIVDEKTVVHTSQTVPDPOEGSTPLMGQAGTP 240  
DB 202 EQHDQKVCEREMQRIITLPSAFTNPTCEIVDEKTVVHTSQTVPDPOEGTTPMLGQAGTP 261  
QY 241 GA 242  
DB 262 GA 263

RESULT 6  
US-09-864-761-38457  
; Sequence 38457, Application US/09864761  
; Patent No. US20020048763A1  
; GENERAL INFORMATION:  
; APPLICANT: Penn, Sharron G.  
; APPLICANT: Rank, David R.  
; APPLICANT: Hanzel, David K.  
; APPLICANT: Chen, Wensheng  
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR  
; FILE REFERENCE: Aescmca-X-1  
; CURRENT APPLICATION NUMBER: US/09/864,761  
; CURRENT FILING DATE: 2001-05-23  
; PRIOR APPLICATION NUMBER: US 60/180,312  
; PRIOR FILING DATE: 2000-02-04  
; PRIOR APPLICATION NUMBER: US 60/207,456  
; PRIOR FILING DATE: 2000-05-26

```
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Annonax Sequence Listing Engine vers. 1.1
; SEQ ID NO 38457
; LENGTH: 66
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AC002073.1
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 2.2
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 2.5
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 2.3
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 1.9
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 2.2
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 2.2
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 2.6
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 2.2
; OTHER INFORMATION: SWISSPROT HIT: P39881, EVALUATE 3.40e+00
; OTHER INFORMATION: EST_HUMAN HIT: BE299406.1, EVALUATE 5.00e-30
; US-09-864-761-38457
Query Match 25.0%; Score 322; DB 9; Length 66;
Best Local Similarity 100.0%; Pred. No. 2e-23;
Matches 66; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 83 TTSSALPAFTTEIQEASGCADEVQVFAPANALPAPSEAAAVQPVIGISQVRMNSKEK 142
Db 1 TTSSALPAFTTEIQEASGCADEVQVFAPANALPAPSEAAAVQPVIGISQVRMNSKEK 60
QY 143 KDLGTL 148
Db 61 KDLGTL 66
RESULT 7
US-09-987-457-18
; Sequence 18, Application US/09987457
; Publication No. US20030013150A1
; GENERAL INFORMATION:
; APPLICANT: Manosroi, Aranya
; APPLICANT: Manosroi, Jiradej
; APPLICANT: Tayapiwatana, Chatchai
```

```
; APPLICANT: Goetz, Friedrich
; APPLICANT: Werner, Rolf-Guenther
; TITLE OF INVENTION: Methods for Large Scale Protein Production in Prokaryotes
; FILE REFERENCE: 0652.2180001
; CURRENT APPLICATION NUMBER: US/09/987,457
; CURRENT FILING DATE: 2001-11-14
; PRIOR APPLICATION NUMBER: 60/268,573
; PRIOR FILING DATE: 2001-02-15
; PRIOR APPLICATION NUMBER: GB 00 27 782.2
; PRIOR FILING DATE: 2000-11-14
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 18
; LENGTH: 527
; TYPE: PRT
; ORGANISM: Homo sapiens (tpa)
; US-09-987-457-18
Query Match 12.0%; Score 154.5; DB 10; Length 527;
Best Local Similarity 39.6%; Pred. No. 4.7e-06;
Matches 36; Conservative 6; Mismatches 36; Indels 13; Gaps 4;
QY 4 CFWDNGHLYREDQTSAPAGLRCNLNWLDAQSLASAPVS-----GAGNHSYCRNPDE 54
Db 92 CYEDQGISYRGWTSTAESGAECTNW--NSSALAKPKYSGRRPDAILRLGLGNHNYCRNPDR 149
QY 55 DPRGPWCYVSGEAGVPEKRPCEDLRCPETTS 85
Db 150 DSK-PWCYVF-KAGKYSSEFCSTPACSEGNS 178
RESULT 8
US-09-987-455-19
; Sequence 19, Application US/09987455
; Publication No. US20030049729A1
; GENERAL INFORMATION:
; APPLICANT: Aranya Manosroi
; APPLICANT: Jiradej Manosroi
; APPLICANT: Chatchai Tayapiwatana
; APPLICANT: Friedrich Goetz
; APPLICANT: Rolf-Guenther Werner
; TITLE OF INVENTION: Methods for Large Scale Production of Recombinant
; TITLE OF INVENTION: DNA-Derived tPA or K2S Molecules
; FILE REFERENCE: 0652.2190001
; CURRENT APPLICATION NUMBER: US/09/987,455
; CURRENT FILING DATE: 2001-11-14
; PRIOR APPLICATION NUMBER: 60/268,574
; PRIOR FILING DATE: 2001-02-15
; PRIOR APPLICATION NUMBER: GB 0027779.8
; PRIOR FILING DATE: 2000-11-14
; NUMBER OF SEQ ID NOS: 25
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 19
; LENGTH: 527
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-09-987-455-19
Query Match 12.0%; Score 154.5; DB 10; Length 527;
Best Local Similarity 39.6%; Pred. No. 4.7e-06;
Matches 36; Conservative 6; Mismatches 36; Indels 13; Gaps 4;
QY 4 CFWDNGHLYREDQTSAPAGLRCNLNWLDAQSLASAPVS-----GAGNHSYCRNPDE 54
Db 92 CYEDQGISYRGWTSTAESGAECTNW--NSSALAKPKYSGRRPDAILRLGLGNHNYCRNPDR 149
QY 55 DPRGPWCYVSGEAGVPEKRPCEDLRCPETTS 85
Db 150 DSK-PWCYVF-KAGKYSSEFCSTPACSEGNS 178
RESULT 9
US-10-360-101-203
```



```
; Sequence 203, Application US/10360101
; Publication No. US20040009550A1
; GENERAL INFORMATION:
; APPLICANT: Moll, Gert N.
; TITLE OF INVENTION: Export and modification of (poly)peptide in the lantibiotic way
; FILE REFERENCE: 2183-5673
; CURRENT APPLICATION NUMBER: US/10/360,101
; CURRENT FILING DATE: 2003-02-07
; PRIOR APPLICATION NUMBER: EP 02077060.8
; PRIOR FILING DATE: 2002-05-24
; NUMBER OF SEQ ID NOS: 309
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 203
; LENGTH: 527
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: sequence of alteplase
US-10-360-101-203

Query Match 12.0%; Score 154.5; DB 15; Length 527;
Best Local Similarity 39.6%; Pred. No. 4.7e-06;
Matches 36; Conservative 6; Mismatches 36; Indels 13; Gaps 4;

QY 4 CFWDNGHLYREDQTSAPAGRLCLNWLDAQSLASAPVS-----GAGNHSYCRNPDE 54
| : | | | | | : | | | | | : | | | | | : | | | | | : | | | | |
Db 92 CYEDQGISYRGTWSTAESGAECTNW--NSSALAKPYSGRRPDAIRLGLGNHNYCRNPDR 149
| : | | | | | : | | | | | : | | | | | : | | | | | : | | | | |

QY 55 DPRGPWCYVSGEAGVPEKPCEDLRCPEPTS 85
| : | | | | | : | | | | | : | | | | | : | | | | | : | | | | |
Db 150 DSK-PWCYVP-KAGKYSSEFCSTPACSEGNS 178
| : | | | | | : | | | | | : | | | | | : | | | | | : | | | | |

RESULT 10
US-09-969-271-7
; Sequence 7, Application US/09969271
; Patent No. US20020098179A1
; GENERAL INFORMATION:
; APPLICANT: Pfizer Inc. (All designated States except GB and EP (GB));
; APPLICANT: Pfizer Limited (GB and EP (GB) only)
; TITLE OF INVENTION: Pharmaceutical Combinations
; FILE REFERENCE: PCS10951APME
; CURRENT APPLICATION NUMBER: US/09/969,271
; CURRENT FILING DATE: 2001-10-01
; PRIOR APPLICATION NUMBER: GB 0025473.0
; PRIOR FILING DATE: 2000-10-17
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 7
; LENGTH: 562
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-969-271-7

Query Match 12.0%; Score 154.5; DB 9; Length 562;
Best Local Similarity 39.6%; Pred. No. 5.1e-06;
Matches 36; Conservative 6; Mismatches 36; Indels 13; Gaps 4;

QY 4 CFWDNGHLYREDQTSAPAGRLCLNWLDAQSLASAPVS-----GAGNHSYCRNPDE 54
| : | | | | | : | | | | | : | | | | | : | | | | | : | | | | |
Db 127 CYEDQGISYRGTWSTAESGAECTNW--NSSALAKPYSGRRPDAIRLGLGNHNYCRNPDR 184
| : | | | | | : | | | | | : | | | | | : | | | | | : | | | | |

QY 55 DPRGPWCYVSGEAGVPEKPCEDLRCPEPTS 85
| : | | | | | : | | | | | : | | | | | : | | | | | : | | | | |
Db 185 DSK-PWCYVP-KAGKYSSEFCSTPACSEGNS 213
| : | | | | | : | | | | | : | | | | | : | | | | | : | | | | |

RESULT 11
US-09-974-298-145
; Sequence 145, Application US/09974298
; Patent No. US20020156263A1
; GENERAL INFORMATION:
```

```
; APPLICANT: Chen, Hwei-Mei
; TITLE OF INVENTION: GENES EXPRESSED IN BREAST CANCER
; FILE REFERENCE: PA-0037 P
; CURRENT APPLICATION NUMBER: US/09/974,298
; CURRENT FILING DATE: 2001-10-04
; PRIOR APPLICATION NUMBER: 60/238,331
; PRIOR FILING DATE: 2000-05-10
; NUMBER OF SEQ ID NOS: 194
; SOFTWARE: PERL Program
; SEQ ID NO 145
; LENGTH: 562
; TYPE: PRT
; ORGANISM: Homo sapiens
; NAME/KEY: misc feature
; OTHER INFORMATION: incyte ID No. US20020156263A1 1001470CD1
US-09-974-298-145

Query Match 12.0%; Score 154.5; DB 9; Length 562;
Best Local Similarity 39.6%; Pred. No. 5.1e-06;
Matches 36; Conservative 6; Mismatches 36; Indels 13; Gaps 4;

QY 4 CFWDNGHLYREDQTSAPAGRLCLNWLDAQSLASAPVS-----GAGNHSYCRNPDE 54
| : | | | | | : | | | | | : | | | | | : | | | | | : | | | | |
Db 127 CYEDQGISYRGTWSTAESGAECTNW--NSSALAKPYSGRRPDAIRLGLGNHNYCRNPDR 184
| : | | | | | : | | | | | : | | | | | : | | | | | : | | | | |

QY 55 DPRGPWCYVSGEAGVPEKPCEDLRCPEPTS 85
| : | | | | | : | | | | | : | | | | | : | | | | | : | | | | |
Db 185 DSK-PWCYVP-KAGKYSSEFCSTPACSEGNS 213
| : | | | | | : | | | | | : | | | | | : | | | | | : | | | | |

RESULT 12
US-10-411-037-26
; Sequence 26, Application US/10411037
; Publication No. US20040043446A1
; GENERAL INFORMATION:
; APPLICANT: Neose Technologies, Inc.
; APPLICANT: Defrees, Shawn
; APPLICANT: Zopf, David
; APPLICANT: Bayer, Robert
; APPLICANT: Hakes, David
; APPLICANT: Chen, Xi
; APPLICANT: Bowe, Caryn
; TITLE OF INVENTION: ALPHA GALACTOSIDASE A: REMODELING AND GLYCOCONJUGATION OF ALPHA
; FILE REFERENCE: 040853-01-5082
; CURRENT APPLICATION NUMBER: US/10/411,037
; CURRENT FILING DATE: 2003-04-09
; PRIOR APPLICATION NUMBER: US 60/328,523
; PRIOR FILING DATE: 2001-10-10
; PRIOR APPLICATION NUMBER: US 60/344,692
; PRIOR FILING DATE: 2001-10-19
; PRIOR APPLICATION NUMBER: US 60/387,292
; PRIOR FILING DATE: 2002-06-07
; PRIOR APPLICATION NUMBER: US 60/391,777
; PRIOR FILING DATE: 2002-06-25
; PRIOR APPLICATION NUMBER: US 60/396,594
; PRIOR FILING DATE: 2002-07-17
; PRIOR APPLICATION NUMBER: US 60/404,249
; PRIOR FILING DATE: 2002-08-16
; PRIOR APPLICATION NUMBER: US 60/407,527
; PRIOR FILING DATE: 2002-08-28
; NUMBER OF SEQ ID NOS: 75
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 26
; LENGTH: 562
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-411-037-26

Query Match 12.0%; Score 154.5; DB 12; Length 562;
Best Local Similarity 39.6%; Pred. No. 5.1e-06;
Matches 36; Conservative 6; Mismatches 36; Indels 13; Gaps 4;
```

```

RESULT 14
US-10-443-701-4
; Sequence 4, Application US/10443701
; Publication No. US20030-99016A1
; GENERAL INFORMATION:
; APPLICANT: Xu, Yuan
; TITLE OF INVENTION: REVERSE-PHASE HPLC ASSAY FOR PLASMINOGEN ACTIVATORS
; FILE REFERENCE: P1788R1
; CURRENT APPLICATION NUMBER: US/10/443,701
; CURRENT FILING DATE: 2003-05-21
; PRIOR APPLICATION NUMBERS: US/09/703,695
; PRIOR FILING DATE: 2000-11-01
; PRIOR APPLICATION NUMBER: US 60/163,607
; PRIOR FILING DATE: 1999-11-04
; NUMBER OF SEQ ID NOS: 4
; SEQ ID NO 4
; LENGTH: 562
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-443-701-4

```

Search completed: March 18, 2004, 13:24:45  
Job time : 44.9704 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 17, 2004, 07:00:39 ; Search time 17.9139 Seconds  
(without alignments)  
697.420 Million cell updates/sec

Title: US-10-057-951-2\_COPY\_22\_263  
Perfect score: 1289  
Sequence: 1 SGGCFWNGHLYREDQTSFA.....PVDPOEGSTPLMGQATPGA 242

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued Patents AA:\*  
1: /cgn2\_6/prodata/2/iaa/5A\_COMB.pep:\*  
2: /cgn2\_6/prodata/2/iaa/5B\_COMB.pep:\*  
3: /cgn2\_6/prodata/2/iaa/6A\_COMB.pep:\*  
4: /cgn2\_6/prodata/2/iaa/6B\_COMB.pep:\*  
5: /cgn2\_6/prodata/2/iaa/PTUS\_COMB.pep:\*  
6: /cgn2\_6/prodata/2/iaa/backfiles1.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1289	100.0	263	US-09-411-977-2	Sequence 2, Appli
2	158.5	12.3	472	US-08-811-949-63	Sequence 63, Appl
3	154.5	12.0	437	US-08-811-949-49	Sequence 49, Appl
4	154.5	12.0	437	US-08-811-949-51	Sequence 51, Appl
5	154.5	12.0	437	US-08-811-949-55	Sequence 55, Appl
6	154.5	12.0	437	US-08-811-949-57	Sequence 57, Appl
7	154.5	12.0	527	US-07-609-5109-16	Sequence 16, Appl
8	154.5	12.0	527	US-08-811-949-39	Sequence 39, Appl
9	154.5	12.0	527	PCT-US91-01025A-2	Sequence 2, Appli
10	154.5	12.0	527	5185259-8	Patent No. 5185259
11	154.5	12.0	527	5520913-1	Patent No. 5520913
12	154.5	12.0	546	5200340-6	Patent No. 5200340
13	154.5	12.0	562	US-08-811-949-43	Sequence 43, Appl
14	154.5	12.0	562	US-08-560-098A-50	Sequence 50, Appl
15	154.5	12.0	562	US-08-883-795A-38	Sequence 38, Appl
16	154.5	12.0	562	US-09-703-695A-4	Sequence 4, Appli
17	154.5	12.0	562	5185259-3	Patent No. 5185259
18	154.5	12.0	562	5200340-2	Patent No. 5200340
19	154.5	12.0	562	5344773-2	Patent No. 5344773
20	148.5	11.5	485	US-08-811-949-2	Sequence 2, Appli
21	146.5	11.4	655	US-08-148-910-12	Sequence 12, Appl
22	146.5	11.4	655	US-08-448-937A-12	Sequence 12, Appl
23	142	11.0	160	US-08-612-788-35	Sequence 35, Appl
24	142	11.0	160	US-08-066-028-35	Sequence 35, Appl
25	142	11.0	160	US-09-335-325-35	Sequence 35, Appl
26	142	11.0	250	US-08-612-788-30	Sequence 30, Appl
27	142	11.0	250	US-09-066-028-30	Sequence 30, Appl

28	142	11.0	250	4	US-09-335-325-30	Sequence 30, Appl
29	142	11.0	339	1	US-08-248-629A-3	Sequence 3, Appli
30	142	11.0	339	1	US-08-451-932-3	Sequence 3, Appli
31	142	11.0	339	1	US-08-452-260-3	Sequence 3, Appli
32	142	11.0	339	1	US-08-326-785-3	Sequence 3, Appli
33	142	11.0	339	2	US-08-612-788-3	Sequence 3, Appli
34	142	11.0	339	2	US-08-605-598B-3	Sequence 3, Appli
35	142	11.0	339	2	US-08-429-743-3	Sequence 3, Appli
36	142	11.0	339	2	US-08-866-735-3	Sequence 3, Appli
37	142	11.0	339	3	US-09-066-028-3	Sequence 3, Appli
38	142	11.0	339	4	US-09-335-325-3	Sequence 3, Appli
39	142	11.0	339	5	PCT-US95-05107-3	Sequence 3, Appli
40	142	11.0	352	2	US-08-812-788-40	Sequence 40, Appl
41	142	11.0	352	3	US-09-066-028-40	Sequence 40, Appl
42	142	11.0	352	4	US-09-335-325-40	Sequence 40, Appl
43	142	11.0	374	4	US-09-377-250-3	Sequence 3, Appli
44	142	11.0	375	4	US-09-377-250-2	Sequence 2, Appli
45	142	11.0	378	2	US-08-612-788-42	Sequence 42, Appl

ALIGNMENTS

RESULT 1  
US-09-411-977-2  
; Sequence 2, Application US/09411977  
; Patent No. 6372473  
; GENERAL INFORMATION:  
; APPLICANT: Moore, Paul A.  
; APPLICANT: Ruben, Steven M.  
; APPLICANT: Ebner, Reinhard  
; TITLE OF INVENTION: Tissue Plasminogen Activator-Like Protease  
; FILE REFERENCE: PF378P1  
; CURRENT APPLICATION NUMBER: US/09/411,977  
; EARLIER FILING DATE: 1999-10-04  
; EARLIER APPLICATION NUMBER: 09/084,491  
; EARLIER FILING DATE: 1998-05-27  
; EARLIER APPLICATION NUMBER: 60/048,000  
; EARLIER FILING DATE: 1997-05-28  
; NUMBER OF SEQ ID NOS: 30  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 2  
; LENGTH: 263  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-411-977-2

Query Match	100.0.0%;	Score 1289;	DB 4;	Length 263;
Best Local Similarity	100.0.0%;	Pred. No. 1.2e-128;		
Matches 242;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
Qy	1	SGGCFWNGHLYREDQTSFAPGLRCLNWLDAQGLASAPVSGAGNHSYCRNPDEDPGPW	60	
Db	22	SGGCFWNGHLYREDQTSFAPGLRCLNWLDAQGLASAPVSGAGNHSYCRNPDEDPGPW	81	
Qy	61	CYVSGEAGVPEKRCEDLRCPETTSQALPAFTTEIQEASGPGADEVQVFAPANALPARS	120	
Db	82	CYVSGEAGVPEKRCEDLRCPETTSQALPAFTTEIQEASGPGADEVQVFAPANALPARS	141	
Qy	121	EAAGVQVIGISQVRMNSKEKDLGLTGLVGLITMMVIIAIGAGIILGYSYKRGKDLK	180	
Db	142	EAAGVQVIGISQVRMNSKEKDLGLTGLVGLITMMVIIAIGAGIILGYSYKRGKDLK	201	
Qy	181	EQHDKQVCEREMQITPLSAFTNPTCEIIVDEKTVVHTSQTPTVDPQEGSTPLMGQATGP	240	
Db	202	EQHDKQVCEREMQITPLSAFTNPTCEIIVDEKTVVHTSQTPTVDPQEGSTPLMGQATGP	261	
Qy	241	GA	242	
Db	262	GA	263	

RESULT 2

US-08-811-949-63  
 ; Sequence 63, Application US/08811949  
 ; Patent No. 5840533  
 ; GENERAL INFORMATION:  
 ; APPLICANT: NIWA, MINEO  
 ; APPLICANT: SAITO, YOSHIMASA  
 ; APPLICANT: SASAKI, HITOSHI  
 ; APPLICANT: HAYASHI, MASAKO  
 ; APPLICANT: NOTANI, JOUJI  
 ; APPLICANT: KOBAYASHI, MASAKAZU  
 ; TITLE OF INVENTION: TISSUE PLASMINOGEN ACTIVATOR  
 ; NUMBER OF SEQUENCES: 67  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MAIER & NEUSTADT,  
 ; ADDRESSEE: P.C.  
 ; STREET: 1755 S. JEFFERSON DAVIS HIGHWAY, SUITE 400  
 ; CITY: ARLINGTON  
 ; STATE: VA  
 ; COUNTRY: USA  
 ; ZIP: 22202  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: Patent In Release #1.0, Version #1.30  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/811,949  
 ; FILING DATE: 05-MAR-1997  
 ; CLASSIFICATION: 435  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: OBLON, NORMAN F.  
 ; REGISTRATION NUMBER: 24,618  
 ; REFERENCE/DOCKET NUMBER: 18-966-0  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: 703-413-3000  
 ; TELEFAX: 703-413-2220  
 ; INFORMATION FOR SEQ ID NO: 63:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 472 amino acids  
 ; TYPE: amino acid  
 ; TOPOLOGY: linear  
 ; MOLECULE TYPE: protein  
 ; US-08-811-949-63

Query Match 12.3%; Score 158.5; DB 2; Length 472;  
 Best Local Similarity 40.7%; Pred. No. 3.2e-08;  
 Matches 37; Conservative 5; Mismatches 36; Indels 13; Gaps 4;  
 QY 4 CFWDNGHLYREDQTSPPAGRLCLNWLDAQSGLASAPVS-----GAGNHSYCRNPDE 54  
 DB 37 CYEDQGISYRGTTWTAESGAECTNW--NSSALAKPYSGRRPDPPIRLGLGNHNYCRNPDR 94  
 QY 55 DPRGPWCYVSGEAGVPEKRCEDLRCPETTS 85  
 DB 95 DSK-PWCYVF-KAGKYSSEFCSTPACSEGNS 123

RESULT 3  
 US-08-811-949-49  
 ; Sequence 49, Application US/08811949  
 ; Patent No. 5840533  
 ; GENERAL INFORMATION:  
 ; APPLICANT: NIWA, MINEO  
 ; APPLICANT: SAITO, YOSHIMASA  
 ; APPLICANT: SASAKI, HITOSHI  
 ; APPLICANT: HAYASHI, MASAKO  
 ; APPLICANT: NOTANI, JOUJI  
 ; APPLICANT: KOBAYASHI, MASAKAZU  
 ; TITLE OF INVENTION: TISSUE PLASMINOGEN ACTIVATOR  
 ; NUMBER OF SEQUENCES: 67  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MAIER & NEUSTADT,  
 ; ADDRESSEE: P.C.

STREET: 1755 S. JEFFERSON DAVIS HIGHWAY, SUITE 400  
 CITY: ARLINGTON  
 STATE: VA  
 COUNTRY: USA  
 ZIP: 22202  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: Patent In Release #1.0, Version #1.30  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/811,949  
 FILING DATE: 05-MAR-1997  
 CLASSIFICATION: 435  
 ATTORNEY/AGENT INFORMATION:  
 NAME: OBLON, NORMAN F.  
 REGISTRATION NUMBER: 24,618  
 REFERENCE/DOCKET NUMBER: 18-966-0  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: 703-413-3000  
 TELEFAX: 703-413-2220  
 INFORMATION FOR SEQ ID NO: 49:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 437 amino acids  
 TYPE: amino acid  
 TOPOLOGY: linear  
 MOLECULE TYPE: protein  
 US-08-811-949-49

Query Match 12.0%; Score 154.5; DB 2; Length 437;  
 Best Local Similarity 39.6%; Pred. No. 7.5e-08;  
 Matches 36; Conservative 6; Mismatches 36; Indels 13; Gaps 4;  
 QY 4 CFWDNGHLYREDQTSPPAGRLCLNWLDAQSGLASAPVS-----GAGNHSYCRNPDE 54  
 DB 2 CYEDQGISYRGTTWTAESGAECTNW--NSSALAKPYSGRRPDPPIRLGLGNHNYCRNPDR 59  
 QY 55 DPRGPWCYVSGEAGVPEKRCEDLRCPETTS 85  
 DB 60 DSK-PWCYVF-KAGKYSSEFCSTPACSEGNS 88

RESULT 4  
 US-08-811-949-51  
 ; Sequence 51, Application US/08811949  
 ; Patent No. 5840533  
 ; GENERAL INFORMATION:  
 ; APPLICANT: NIWA, MINEO  
 ; APPLICANT: SAITO, YOSHIMASA  
 ; APPLICANT: SASAKI, HITOSHI  
 ; APPLICANT: HAYASHI, MASAKO  
 ; APPLICANT: NOTANI, JOUJI  
 ; APPLICANT: KOBAYASHI, MASAKAZU  
 ; TITLE OF INVENTION: TISSUE PLASMINOGEN ACTIVATOR  
 ; NUMBER OF SEQUENCES: 67  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MAIER & NEUSTADT,  
 ; ADDRESSEE: P.C.  
 ; STREET: 1755 S. JEFFERSON DAVIS HIGHWAY, SUITE 400  
 ; CITY: ARLINGTON  
 ; STATE: VA  
 ; COUNTRY: USA  
 ; ZIP: 22202  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: Patent In Release #1.0, Version #1.30  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/811,949  
 ; FILING DATE: 05-MAR-1997  
 ; CLASSIFICATION: 435  
 ; ATTORNEY/AGENT INFORMATION:

NAME: OBLON, NORMAN F.  
REGISTRATION NUMBER: 24,618  
REFERENCE/DOCKET NUMBER: 18-966-0  
TELEPHONE: 703-413-3000  
TELEFAX: 703-413-2220  
INFORMATION FOR SEQ ID NO: 51:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 437 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-811-949-51

Query Match 12.0%; Score 154.5; DB 2; Length 437;  
Best Local Similarity 39.6%; Pred. No. 7.5e-08;  
Matches 36; Conservative 6; Mismatches 36; Indels 13; Gaps 4;

QY 4 CFWDNGHLYREDQTSAPAGLCRLNWLDAQSLASAPVS-----GAGNHSYCRNPDE 54  
Db 2 CYEDQGISYRGWTSTAESGAECTNW--NSSALAKPYSGRRPDAILRLGLGNHNYCRNPDR 59  
QY 55 DPRGWCYVSGEAGVPEKPCEDLRCPETTS 85  
Db 60 DSK-PWCYVF-KAGKYSSEFCSTPACSEGS 88

## RESULT 5

US-08-811-949-55  
Sequence 55, Application US/08811949  
Patent No. 5840533  
GENERAL INFORMATION:  
APPLICANT: NIWA, MINEO  
APPLICANT: SAITO, YOSHIMASA  
APPLICANT: SASAKI, HITOSHI  
APPLICANT: HAYASHI, MASAKO  
APPLICANT: NOTANI, JOUJI  
APPLICANT: KOBAYASHI, MASAKAZU  
TITLE OF INVENTION: TISSUE PLASMINOGEN ACTIVATOR  
NUMBER OF SEQUENCES: 67  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MAIER & NEUSTADT,  
ADDRESS: P.C.  
STREET: 1755 S. JEFFERSON DAVIS HIGHWAY, SUITE 400  
CITY: ARLINGTON  
STATE: VA  
COUNTRY: USA  
ZIP: 22202  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/811,949  
FILING DATE: 05-MAR-1997  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: OBLON, NORMAN F.  
REGISTRATION NUMBER: 24,618  
REFERENCE/DOCKET NUMBER: 18-966-0  
TELEPHONE: 703-413-3000  
TELEFAX: 703-413-2220  
INFORMATION FOR SEQ ID NO: 55:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 437 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-811-949-55

Query Match 12.0%; Score 154.5; DB 2; Length 437;

Best Local Similarity 39.6%; Pred. No. 7.5e-08;  
Matches 36; Conservative 6; Mismatches 36; Indels 13; Gaps 4;  
QY 4 CFWDNGHLYREDQTSAPAGLCRLNWLDAQSLASAPVS-----GAGNHSYCRNPDE 54  
Db 2 CYEDQGISYRGWTSTAESGAECTNW--NSSALAKPYSGRRPDAILRLGLGNHNYCRNPDR 59  
QY 55 DPRGWCYVSGEAGVPEKPCEDLRCPETTS 85  
Db 60 DSK-PWCYVF-KAGKYSSEFCSTPACSEGS 88

## RESULT 6

US-08-811-949-57  
Sequence 57, Application US/08811949  
Patent No. 5840533  
GENERAL INFORMATION:  
APPLICANT: NIWA, MINEO  
APPLICANT: SAITO, YOSHIMASA  
APPLICANT: SASAKI, HITOSHI  
APPLICANT: HAYASHI, MASAKO  
APPLICANT: NOTANI, JOUJI  
APPLICANT: KOBAYASHI, MASAKAZU  
TITLE OF INVENTION: TISSUE PLASMINOGEN ACTIVATOR  
NUMBER OF SEQUENCES: 67  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MAIER & NEUSTADT,  
ADDRESS: P.C.  
STREET: 1755 S. JEFFERSON DAVIS HIGHWAY, SUITE 400  
CITY: ARLINGTON  
STATE: VA  
COUNTRY: USA  
ZIP: 22202  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/811,949  
FILING DATE: 05-MAR-1997  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: OBLON, NORMAN F.  
REGISTRATION NUMBER: 24,618  
REFERENCE/DOCKET NUMBER: 18-966-0  
TELEPHONE: 703-413-3000  
TELEFAX: 703-413-2220  
INFORMATION FOR SEQ ID NO: 57:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 437 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-811-949-57

Query Match 12.0%; Score 154.5; DB 2; Length 437;  
Best Local Similarity 39.6%; Pred. No. 7.5e-08;  
Matches 36; Conservative 6; Mismatches 36; Indels 13; Gaps 4;

QY 4 CFWDNGHLYREDQTSAPAGLCRLNWLDAQSLASAPVS-----GAGNHSYCRNPDE 54  
Db 2 CYEDQGISYRGWTSTAESGAECTNW--NSSALAKPYSGRRPDAILRLGLGNHNYCRNPDR 59  
QY 55 DPRGWCYVSGEAGVPEKPCEDLRCPETTS 85  
Db 60 DSK-PWCYVF-KAGKYSSEFCSTPACSEGS 88

## RESULT 7

US-07-609-510B-16  
Sequence 16, Application US/07609510B

```

; Patent No. 5326700
;
; GENERAL INFORMATION:
;
; APPLICANT: Berg et al.
;
; TITLE OF INVENTION: Method for Altering Post-Translational Processing of Tissue PL
;
; NUMBER OF SEQUENCES: 16
;
; CORRESPONDENCE ADDRESS:
;
; ADDRESSEE: Eli Lilly and Company
;
; STREET: Lilly Corporate Center
;
; CITY: Indianapolis
;
; STATE: IN.
;
; COUNTRY: U.S.A.
;
; ZIP: 46285
;
; COMPUTER READABLE FORM:
;
; MEDIUM TYPE: Diskette, 3.50 inch, 1.0 Mb storage
;
; COMPUTER: Macintosh
;
; OPERATING SYSTEM: Macintosh
;
; SOFTWARE: Microsoft Word
;
; CURRENT APPLICATION DATA:
;
; APPLICATION NUMBER: US/07/609,510B
;
; FILING DATE: 19901106
;
; CLASSIFICATION: 435
;
; INFORMATION FOR SEQ ID NO: 16:
;
; SEQUENCE CHARACTERISTICS:
;
; LENGTH: 527 amino acids
;
; TYPE: AMINO ACID
;
; STRANDEDNESS: single
;
; TOPOLOGY: linear
;
; MOLECULE TYPE: protein
;
; US-07-609-510B-16

```

```

Query Match      12.0%; Score 154.5; DB 1; Length 527;
Best Local Similarity 39.6%; Pred.No. 9.9e-08;
Matches 36; Conservative 6; Mismatches 36; Indels 13; Gaps 4;

QY      4  CFWDNGHLYREDQTSAPGLRCLNWLDAQGLASAPVS-----GAGNSHCYCRNPDE 54
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      92  CYEDQGISYRGWTWTAESGAECTNW--NSSALAKQPSYGRPDARLGLGNHNYCRNPDR 149

QY      55  DPRGPWCYVSGEAGVFEKRRPCDLERCPEETS 85
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      150  DGK-PWCYVF-KAGKYSSEFCSTPACGEGNS 178

```

RESULT 8  
US-08-811-949-39  
; Sequence 39, Application US/08811949  
; Patent No. 5840533  
; GENERAL INFORMATION:  
; APPLICANT: NIWA, MINEO  
; APPLICANT: SAITO, YOSHIMASA  
; APPLICANT: SASAKI, HITOSHI  
; APPLICANT: HAYASHI, MASAKO  
; APPLICANT: NOTANI, JOUJI  
; APPLICANT: KOBAYASHI, MASAKAZU  
; TITLE OF INVENTION: TISSUE PLASMINOGEN ACTIVATOR  
; NUMBER OF SEQUENCES: 67  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MAIER & NEUSTADT,  
; ADDRESSEE: P.C.  
; STREET: 1755 S. JEFFERSON DAVIS HIGHWAY, SUITE 400  
; CITY: ARLINGTON  
; STATE: VA  
; COUNTRY: USA  
; ZIP: 22202  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/811,949  
; FILING DATE: 05-MAR-1997  
; CLASSIFICATION: 435

```

; ATTORNEY/AGENT INFORMATION:
; NAME: OBLON, NORMAN F.
; REGISTRATION NUMBER: 24,618
; REFERENCE/DOCKET NUMBER: 18-966-0
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 703-413-3000
; TELEFAX: 703-413-2220
; INFORMATION FOR SEQ ID NO: 39:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 527 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-811-949-39

Query Match          12.0%; Score 154.5; DB 2; Length 527;
Best Local Similarity 39.8%; Pred. No. 9.9e-08;
Matches 36; Conservative 6; Mismatches 36; Indels 13; Gaps 4;

QY      4 C F D N G H L Y R E D T S P A P G L R C L N W L D A Q G L S A P V S -----GAGNHYSYCRNPDE 54
         : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      92 C Y E D Q G I S Y G T W S T A E S G A E C T I N W - N S S A L A K Q P Y S G R R P D A I R L G L G N H Y C R N P D R 149

QY      55 D P R G P W C Y V S G E A G V P E K R F C E D L R C P E T T S 85
         : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db     150 D S K - P W C Y V F - K A G K Y S S E F C S T P A G S E G N S 178

```

```

RESULT 9
PCT-US91-01025A-2
: Sequence 2, Application PC/TUS9101025A
:
: GENERAL INFORMATION:
:
: APPLICANT: Genentech, Inc.
:
: TITLE OF INVENTION: Tissue Plasminogen Activator Having Fibrin
:
: TITLE OF INVENTION: Specific Properties
:
: NUMBER OF SEQUENCES: 2
:
: CORRESPONDENCE ADDRESS:
:
: ADDRESSEE: Genentech, Inc.
:
: STREET: 460 Point San Bruno Blvd
:
: CITY: South San Francisco
:
: STATE: California
:
: COUNTRY: USA
:
: ZIP: 94080

```

COMPUTER READABLE FORM:  
MEDIUM TYPE: 5.25 inch, 360 KB floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: patin (Genentech)  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: PCT/US91/01025A  
FILING DATE: 19910214  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 07/486,657  
FILING DATE: 1 March 1990  
ATTORNEY/AGENT INFORMATION:  
NAME: Hasak, Janet E.  
REGISTRATION NUMBER: 28,616  
REFERENCE/DOCKET NUMBER: 454P2  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 415/266-1896  
TELEFAX: 415/952-9881  
TELEX: 910/371-7168  
INFORMATION FOR SEQ ID NO.: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 527 amino acids  
TYPE: AMINO ACID  
TOPOLOGY: linear  
PCT-US91-01025A-2

Query Match 12.0%; Score 154.5; DB 5; Length 527;  
Best Local Similarity 39.6%; Pred. No. 9.9e-08;  
Matches 36; Conservative 6; Mismatches 36; Indels 13; Gaps 4;



TELEFAX: 703-413-2220  
; INFORMATION FOR SEQ ID NO: 43:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 562 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-811-949-43  
  
Query Match 12.0%; Score 154.5; DB 2; Length 562;  
Best Local Similarity 39.6%; Pred. No. 1.1e-07;  
Matches 36; Conservative 6; Mismatches 36; Indels 13; Gaps 4;  
  
QY 4 CFWDNGHLYREDQTSAPGLRCLNWLDAQSLASAPVS-----GAGNHSYCRNPDE 54  
Db 127 CYEDQGISYRGTWSTAESGAECTNW--NSSALAQPKYSGRRPDAIRLGLGNHNYCRNPDR 184  
  
QY 55 DPRGPWCYVSGEAGVPEKPCEDLRCPETTS 85  
Db 185 DSK-PWCYVF-KAGKYSSEFCSTPACSEGS 213  
  
RESULT 14  
US-08-560-098A-50  
; Sequence 50, Application US/08560098A  
; Patent No. 5976841  
; GENERAL INFORMATION:  
; APPLICANT: WENDET, Stephan  
; APPLICANT: HEINZEL-WIELAND, Regina  
; APPLICANT: STEFFENS, Gerd Josef  
; TITLE OF INVENTION: Proteins having fibrinolytic and  
; TITLE OF INVENTION: Coagulation-inhibiting Properties  
; NUMBER OF SEQUENCES: 60  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Evenson, McKeown, Edwards & Lenahan  
; STREET: 1200 G Street, N.W., Suite 700  
; CITY: Washington  
; STATE: DC  
; COUNTRY: USA  
; ZIP: 20005  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/560,098A  
; FILING DATE: 17-NOV-1995  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: P 44 40 892.7  
; FILING DATE: 17-NOV-1994  
; ATTORNEY/AGENT INFORMATION:  
; NAME: EVANS, Joseph D.  
; REGISTRATION NUMBER: 26,269  
; REFERENCE/DOCKET NUMBER: 148/42448  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (202) 628-8800  
; TELEFAX: (202) 628-8844  
; INFORMATION FOR SEQ ID NO: 50:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 562 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-560-098A-50

Query Match 12.0%; Score 154.5; DB 2; Length 562;  
Best Local Similarity 39.6%; Pred. No. 1.1e-07;  
Matches 36; Conservative 6; Mismatches 36; Indels 13; Gaps 4;  
  
QY 4 CFWDNGHLYREDQTSAPGLRCLNWLDAQSLASAPVS-----GAGNHSYCRNPDE 54  
Db 127 CYEDQGISYRGTWSTAESGAECTNW--NSSALAQPKYSGRRPDAIRLGLGNHNYCRNPDR 184

Db 127 CYEDQGISYRGTWSTAESGAECTNW--NSSALAQPKYSGRRPDAIRLGLGNHNYCRNPDR 184  
  
QY 55 DPRGPWCYVSGEAGVPEKPCEDLRCPETTS 85  
Db 185 DSK-PWCYVF-KAGKYSSEFCSTPACSEGS 213  
  
RESULT 15  
US-08-883-795A-38  
; Sequence 38, Application US/08883795A  
; Patent No. 5985607  
; GENERAL INFORMATION:  
; APPLICANT: Delcuve, Genevieve  
; APPLICANT: Awang, Gregor  
; TITLE OF INVENTION: Recombinant DNA Molecules and Expression  
; TITLE OF INVENTION: Vectors for Tissue Plasminogen Activator  
; NUMBER OF SEQUENCES: 39  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: BERESKIN & PARR  
; STREET: 40 King Street West  
; CITY: Toronto  
; STATE: Ontario  
; COUNTRY: Canada  
; ZIP: M5H 3Y2  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/883,795A  
; FILING DATE: 27-JUN-1997  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Gravelle, Micheline  
; REGISTRATION NUMBER: 40,261  
; REFERENCE/DOCKET NUMBER: 7841-062  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (416) 364-7311  
; TELEFAX: (416) 364-1398  
; INFORMATION FOR SEQ ID NO: 38:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 562 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein (tPA)  
US-08-883-795A-38

Query Match 12.0%; Score 154.5; DB 2; Length 562;  
Best Local Similarity 39.6%; Pred. No. 1.1e-07;  
Matches 36; Conservative 6; Mismatches 36; Indels 13; Gaps 4;  
  
QY 4 CFWDNGHLYREDQTSAPGLRCLNWLDAQSLASAPVS-----GAGNHSYCRNPDE 54  
Db 127 CYEDQGISYRGTWSTAESGAECTNW--NSSALAQPKYSGRRPDAIRLGLGNHNYCRNPDR 184  
  
QY 55 DPRGPWCYVSGEAGVPEKPCEDLRCPETTS 85  
Db 185 DSK-PWCYVF-KAGKYSSEFCSTPACSEGS 213

Search completed: March 17, 2004, 07:09:07  
Job time : 18.9139 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 17, 2004, 06:59:28 ; Search time 4.19919 Seconds  
(without alignments)  
1374.429 Million cell updates/sec

Title: US-10-057-951-2\_COPY\_25\_84  
Perfect score: 357  
Sequence: 1 CFWDNGHLYREDQTSAPGL.....GNHSYCRNPDBDPRGWCYV 60

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : PIR 78: \*  
1: Pirl: \*  
2: Pirl: \*  
3: Pirl: \*  
4: Pirl: \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	146	40.9	562	1 UKHUT	t-plasminogen acti
2	141	39.5	559	1 A35029	t-plasminogen acti
3	141	39.5	655	1 A46888	hepatocyte growth
4	137	38.4	291	2 I38098	t-plasminogen acti
5	136	38.1	559	1 A29941	t-plasminogen acti
6	134	37.5	431	2 JS0599	t-plasminogen acti
7	134	37.5	477	1 A34369	t-plasminogen acti
8	134	37.5	477	2 JS0598	t-plasminogen acti
9	132	37.0	433	1 UKBAY	u-plasminogen acti
10	131	36.7	431	1 UKHU	u-plasminogen acti
11	130	36.4	442	1 UKPG	u-plasminogen acti
12	128	35.9	716	1 JC5061	macrophage-stimula
13	127	35.6	433	1 JN0560	u-plasminogen acti
14	126	35.3	394	2 JS0600	t-plasminogen acti
15	126	35.3	716	1 A40332	macrophage-stimula
16	123	34.5	169	2 A40322	Plasmin (EC 3.4.21
17	121	33.9	810	1 PLHU	Plasmin (EC 3.4.21
18	119	33.3	433	1 UKMS	u-plasminogen acti
19	119	33.3	728	1 JH0579	hepatocyte growth
20	118	33.1	432	1 S18932	u-plasminogen acti
21	118	33.1	728	1 A60185	hepatocyte growth
22	118	33.1	810	2 T46260	plasmin (EC 3.4.21
23	115	32.2	477	2 JS0597	t-plasminogen acti
24	113.5	31.8	790	1 PLPG	plasmin (EC 3.4.21
25	113.5	31.8	4548	1 SC0657	apoptein(a) (EC
26	113	31.7	820	2 B30848	plasmin (EC 3.4.21
27	112	31.4	728	1 A35644	hepatocyte growth
28	111.5	31.2	603	2 S28941	coagulation factor
29	111	31.1	812	1 PLBO	plasmin (EC 3.4.21

30	109	30.5	685	1 A48289	neurotrophic recep
31	109	30.5	711	1 A47136	macrophage-stimula
32	108	30.3	625	1 TBBO	thrombin (EC 3.4.21
33	107.5	30.1	123	2 C61545	plasmin (EC 3.4.21
34	106.5	29.8	434	1 A35005	u-plasminogen acti
35	105.5	29.6	460	2 B61545	plasmin (EC 3.4.21
36	105	29.4	622	1 TBHU	thrombin (EC 3.4.21
37	105	29.4	710	1 I51283	hepatocyte growth
38	104.5	29.3	812	1 PLMS	plasmin (EC 3.4.21
39	104	29.1	560	1 JC4795	plasma hyaluronan-
40	101.5	28.4	120	2 B61545	plasmin (EC 3.4.21
41	100.5	28.2	455	2 A61545	plasmin (EC 3.4.21
42	98.5	27.6	593	2 S45281	coagulation factor
43	97	27.2	617	2 S10511	thrombin (EC 3.4.21
44	95.5	26.8	89	2 A60140	plasmin (EC 3.4.21
45	95	26.6	618	2 A35827	thrombin (EC 3.4.21

ALIGNMENTS

RESULT 1

UKHUT

t-plasminogen activator (EC 3.4.21.68) precursor [validated] - human

N:Alternate names: t-PA; tissue plasminogen activator

C:Species: Homo sapiens (man)

C>Date: 14-Nov-1983 #sequence revision 14-Nov-1983 #text change 08-Dec-2000

C:Accession: A94004; A23529; J0562; A93293; S02125; A91343; A93951; A91322; A54645; I6

R:NY, T.; Elgh, F.; Lund, B.

A:Title: The structure of the human tissue-type plasminogen activator gene: correlation

A:Reference number: A94004; MUID:84298137; PMID:6089198

A:Accession: A94004

A:Molecule type: DNA

A:Residues: 1-562 <NT>

A:Cross-references: GB:J00141

A>Note: The codon given for residue 93 (ACC) is inconsistent with the authors' translat

R:Friezeher Degen, S.J.; Rajput, B.; Reich, E.

J. Biol. Chem. 261, 6972-6985, 1986

A:Title: The human tissue plasminogen activator gene.

A:Reference number: A23529; MUID:86196143; PMID:3009482

A:Accession: A23529

A:Molecule type: DNA

A:Residues: 1-562 <DSG>

A:Cross-references: GB:K03021; NID:G339817; PIDN:AAA98809.1; PID:G339818

R:Itagaki, Y.; Yasuda, H.; Morinaga, T.; Mitsuda, S.; Higashio, K.

Agric. Biol. Chem. 55, 1225-1232, 1991

A:Title: Purification and characterization of tissue plasminogen activator secreted by

A:Reference number: J0562; MUID:91291340; PMID:1368681

A:Accession: J0562

A:Molecule type: mRNA

A:Residues: 31-562 <ITA>

A:Cross-references: DBJ:D01096; NID:G220128; PIDN:BAA00881.1; PID:G441174

A:Experimental source: embryonic lung fibroblast IMR-90 cells

A>Note: part of this sequence, including the amino end of the mature protein, was confi

Ripennica, D.; Holmes, W.E.; Kohr, W.J.; Harkins, R.N.; Vehar, G.A.; Ward, C.A.; Bennet

Nature 301, 214-221, 1983

A:Title: Cloning and expression of human tissue-type plasminogen activator cDNA in Esch

A:Reference number: A93293; MUID:83115262; PMID:6337343

A:Accession: A93293

A:Molecule type: mRNA

A:Residues: 1-562 <PEN>

A:Cross-references: GB:L00141

A:Experimental source: melanoma cells

R:Sasaki, H.; Saito, Y.; Hayashi, M.; Otsuka, K.; Niwa, M.

Nucleic Acids Res. 16, 5695, 1988

A:Title: Nucleotide sequence of the tissue-type plasminogen activator cDNA from human f

A:Reference number: S02125; MUID:88262579; PMID:3133640

A:Accession: S02125

A>Status: translation not shown

A:Molecule type: mRNA

A:Residues: 1-562 <SAS>

A:Cross-references: EMBL:X07393; NID:G37243; PIDN:CAA30302.1; PID:G37244

A;Experimental source: fetal lung cells  
R;Kagitani, H.; Tagawa, M.; Hatanaka, K.; Ikari, T.; Saito, A.; Bando, H.; Okada, K.; Ma  
FEBS Lett. 189, 145-149, 1985  
A;Title: Expression in *Escherichia coli* of finger-domain lacking tissue-type plasminogen  
A;Reference number: A91343; MUID:85285620; PMID:3896853  
A;Accession: A91343  
A;Molecule type: mRNA  
A;Residues: 1-38, 'G', 86-433, 'E', 435-562 <KAG>  
A;Experimental source: Detroit 562 cells; ATCC 138  
R;Edlund, T.; Ny, T.; Ranby, M.; Heden, L.O.; Palm, G.; Holmgren, E.; Josephson, S.  
Proc. Natl. Acad. Sci. U.S.A. 80, 349-352, 1983  
A;Title: Isolation of cDNA sequences coding for a part of human tissue plasminogen activator  
A;Reference number: A93951; MUID:83169656; PMID:6572897  
A;Accession: A93951  
A;Molecule type: mRNA  
A;Residues: 251-358 <EDL>  
A;Experimental source: melanoma cells  
R;Pohl, G.; Kallstrom, M.; Bergsdorf, N.; Wallen, P.; Jornvall, H.  
Biochemistry 23, 3701-3707, 1984  
A;Title: Tissue plasminogen activator: peptide analyses confirm an indirectly derived amino acid  
differences.  
A;Reference number: A90488; MUID:85000468; PMID:6433976  
A;Contents: annotation; melanoma cells; partial sequence of residues 36-562, active and  
R;Pohl, G.; Kaplan, L.; Einarsson, M.; Wallen, P.; Jornvall, H.  
FEBS Lett. 168, 29-32, 1984  
A;Title: Differences between uterine and melanoma forms of tissue plasminogen activator.  
A;Reference number: A91322; MUID:84158956; PMID:6538514  
A;Accession: A91322  
A;Molecule type: protein  
A;Residues: 33-45/311-320 <POH>  
A;Experimental source: uterus  
A;Note: in the uterus, cleavage of the activation peptide may also occur after 38-Gln  
R;van Zonneveld, A.J.; Veerman, H.; Pannekoek, H.  
J. Biol. Chem. 261, 14214-14218, 1986  
A;Reference number: A37567; MUID:87033611; PMID:3021732  
A;Contents: annotation; fibrin binding site  
R;Verheijen, J.H.; Caspers, M.P.M.; Chang, G.T.G.; de Munk, G.A.W.; Pourwels, P.H.; Engen  
EMBO J. 5, 3525-3530, 1986  
A;Title: Involvement of finger domain and kringle 2 domain of tissue-type plasminogen activator  
A;Reference number: A37568; MUID:87161761; PMID:3030730  
A;Contents: annotation; fibrin binding site  
R;Dodd, I.; Nunn, B.; Robinson, J.H.  
Thromb. Haemost. 59, 523-528, 1988  
A;Title: Isolation, identification and pharmacokinetic properties of human tissue-type plasminogen  
activator (t-PA) and its recombinant forms.  
A;Reference number: A60902; MUID:89044681; PMID:3142086  
R;Harris, T.J.R.; Patel, T.; Marston, F.A.O.; Little, S.; Entage, J.S.; Opdenakker, G.;  
Mol. Biol. Med. 3, 275-292, 1986  
A;Title: Cloning of cDNA coding for human tissue-type plasminogen activator and its expression  
in *Escherichia coli*.  
A;Reference number: A54645; MUID:86284200; PMID:3090401  
A;Accession: A54645  
A;Molecule type: mRNA  
A;Residues: 1-562 <HAR>  
A;Cross-references: GB:M15518; NID:G190031; PIDN:AAA60111.1; PID:G190032  
A;Note: parts of this sequence were confirmed by peptide sequencing  
R;Reddy, V.B.; Garrazone, A.J.; Saeak, H.; Wei, C.  
DNA 6, 461-472, 1987  
A;Title: Expression of human uterine tissue-type plasminogen activator in mouse cells using  
a cDNA expression vector.  
A;Reference number: I60110; MUID:89054470; PMID:2824147  
A;Accession: I60110  
A;Status: translated from GB/EMBL/DBJ  
A;Molecule type: mRNA  
A;Residues: 1-562 <RES>  
A;Cross-references: GB:M18182; NID:G340176; PIDN:AAA36800.1; PID:G340177  
R;Fisher, R.; Waller, E.K.; Grossi, G.; Thompson, D.; Tizard, R.; Schleuning, W.D.  
J. Biol. Chem. 260, 11223-11230, 1985  
A;Title: Isolation and characterization of the human tissue-type plasminogen activator cDNA  
sequence.  
A;Reference number: I55232; MUID:85289338; PMID:3161893  
A;Accession: I55232  
A;Status: preliminary; translated from GB/EMBL/DBJ  
A;Molecule type: DNA  
A;Residues: 1-36 <RE2>  
A;Cross-references: GB:M11890; NID:G339837; PIDN:AAA61213.1; PID:G339839

C;Comment: Cleavage by plasmin or trypsin produces two chains held together by a single  
disulfide bond. I  
C;Comment: t-PA converts plasminogen to plasmin by hydrolyzing a single Arg-Val bond. I  
C;Comment: t-PA binds chain A of fibrin by kringle 2 and the fibronectin type I repeat.  
C;Genetics:  
A;Gene: GDB:PLAT  
A;Cross-references: GDB:119496; OMIM:173370  
A;Map position: 8p12-8p13  
A;Introns: 24/3; 39/1; 85/1; 122/1; 180/2; 211/1; 268/2; 297/1; 362/2; 408/1; 455/3; 51  
A;Superfamily: tissue plasminogen activator; EGF homology; fibronectin type I repeat ho  
C;Keywords: fibrinolysis; glycoprotein; hydrolase; kringle; plasma; serine proteinase  
F;1-23/Domain: signal sequence #status predicted <SIG>  
F;24-32/Domain: propeptide #status predicted <PRO>  
F;33-562/Product: t-plasminogen activator #status experimental <MAT>  
F;33-310/Product: t-plasminogen activator chain A #status experimental <ACH>  
F;41-78/Domain: fibronectin type I repeat homology <IF1>  
F;86-119/Domain: EGF homology <EGF>  
F;127-208/Domain: kringle homology <KR1>  
F;215-296/Domain: kringle homology <KR2>  
F;311-562/Product: t-plasminogen activator chain B #status experimental <BCH>  
F;311-556/Domain: trypsin homology <TRY>  
F;41-71,69-78,86-97,91-108,110-119,127-208,148-190,179-203,215-296,236-278,267-291,299-  
F;112,483/Binding site: carboxylate (Asn) (covalent) #status experimental  
F;219/Binding site: carboxylate (Asn) (covalent) (partial) #status experimental  
F;310-311/Cleavage site: Arg-11e (plasmin, trypsin) #status experimental  
F;357-406/Active site: His Asp #status predicted  
F;513/Active site: Ser #status experimental  
Query Match 40.9%; Score 146; DB 1; Length 562;  
Best Local Similarity 43.5%; Pred. No. 7.6e-09;  
Matches 30; Conservative 5; Mismatches 22; Indels 12; Gaps 3;  
QY 1 CFWDNGHLYREDQTSPPAGPGLNWLDAQSLASAPVS-----GAGNHSYCRNPDE 51  
DB 127 CYEDQGISYGTWSTAEGACTNW--NSSALAKPYSGRRPDALRLGLGNHYCRNPDR 184  
QY 52 DPGPWCYV 60  
DB 185 DSK-PWCYV 192  
RESULT 2  
A35029  
t-plasminogen activator (EC 3.4.21.68) precursor - rat  
C;Species: Rattus norvegicus (Norway rat)  
C;Date: 10-Sep-1999 #sequence\_revision 10-Sep-1999 #text\_change 10-Sep-1999  
C;Accession: A35029; A31597  
R;Feng, P.; Ohlsson, M.; Ny, T.  
J. Biol. Chem. 265, 2022-2027, 1990  
A;Title: The structure of the TATA-less rat tissue-type plasminogen activator gene. Spe  
cial features of the TATA-less rat tissue-type plasminogen activator gene.  
A;Reference number: A35029; MUID:90130448; PMID:2105315  
A;Accession: A35029  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-559 <FEN>  
A;Cross-references: GB:M31197; NID:G207429; PIDN:AAA42261.1; PID:G207431; GB:J05226  
R;Ny, T.; Leonardsson, G.; Hsueh, A.J.W.  
DNA 7, 671-677, 1988  
A;Title: Cloning and characterization of a cDNA for rat tissue-type plasminogen activat  
ion.  
A;Reference number: A31597; MUID:89170114; PMID:3148445  
A;Accession: A31597  
A;Molecule type: mRNA  
A;Residues: 1-379, 'K', 381-559 <NVT>  
A;Cross-references: GB:M32697; NID:G530159; PIDN:AAA41812.1; PID:G530160  
C;Superfamily: tissue plasminogen activator; EGF homology; fibronectin type I repeat hc  
C;Keywords: fibrinolysis; glycoprotein; hydrolase; kringle; serine proteinase  
F;1-17/Domain: signal sequence #status predicted <SIG>  
F;18-29/Domain: propeptide #status predicted <PRO>  
F;30-559/Product: t-plasminogen activator #status predicted <ACH>  
F;30-308/Product: t-plasminogen activator chain A #status predicted <ACH>  
F;38-75/Domain: fibronectin type I repeat homology <IF1>  
F;83-116/Domain: EGF homology <EGF>  
F;124-205/Domain: kringle homology <KR1>  
F;213-294/Domain: kringle homology <KR2>

F;309-559/Product: t-plasminogen activator chain B #status predicted <BCH>  
F;309-553/Domain: trypsin homology <TRY>  
F;38-66,66-75,83-94,88-105,107-116,124-205,145-187,176-200,213-294,234-276,265-289,297-349,481/Binding site: carbohydrate (Asn) #status predicted  
F;308-309/Cleavage site: Arg-Ile (plasmin, trypsin) #status predicted  
F;355,404,510/Active site: His, Asp, Ser #status predicted

Query Match 39.5%; Score 141; DB 1; Length 559;  
Best Local Similarity 42.0%; Pred. No. 2.8e-08;  
Matches 29; Conservative 6; Mismatches 22; Indels 12; Gaps 3;

QY 1 CFWNQHLHYREDQTSAPAGRLUWLDAQGLASAPVS-----GAGNHSCRNPD 51  
DB 124 CFEGQGITRGTVTAENGAEINW--NSSALSKPYSAARRPNAIKLGLGNHNYCRNPD 181  
QY 52 DPGPCWCV 60  
DB 182 DVK-PWCYV 189

RESULT 3  
A46688  
hepatocyte growth factor activator (EC 3.4.21.-) precursor [validated] - human  
C;Species: Homo sapiens (man)  
C;Date: 21-Sep-1993 #sequence\_revision 25-Aug-1995 #text\_change 08-Dec-2000  
C;Accession: A46688  
R;Miyazawa, K.; Shimomura, T.; Kitamura, A.; Kondo, J.; Morimoto, Y.; Kitamura, N.  
J. Biol. Chem. 269, 10024-10028, 1993  
A;Title: Molecular cloning and sequence analysis of the cDNA for a human serine protease  
d coagulation factor XII.  
A;Reference number: A46688; MUID:93252878; PMID:7683665  
A;Accession: A46688  
A;Molecule type: mRNA  
A;Residues: 1-655 <MY>  
A;Cross-references: DBJ:U14012; NID:g219680; PIDN:BA03113.1; PID:g219681  
A;Experimental source: liver (NRB); serum (protein)  
A;Note: sequence extracted from NCBI backbone (NCBI:131227, NCBI:131228)  
A;Note: parts of the sequence, including the amino ends of the heavy and light chains, C;Genetics:  
A;Gene: GDB:HGFAC; HGFA: HGFAP  
A;Cross-references: GDB:9954514  
A;Map position: 4p16-4p16  
C;Function:  
A;Description: activates hepatocyte growth factor by specific proteolytic cleavage  
A;Pathway: tissue repair and regeneration  
C;Superfamily: coagulation factor XII; EGF homology; fibronectin type I repeat homology;  
C;Keywords: glycoprotein; hydrolase; kringle; liver; plasma; serine proteinase  
F;1-34/Domain: signal sequence #status predicted <SIG>  
F;108-148/Domain: fibronectin type II repeat homology <1F2>  
F;164-197/Domain: EGF homology <EG1>  
F;202-237/Domain: fibronectin type I repeat homology <1F1>  
F;245-278/Domain: EGF homology <EG2>  
F;286-367/Domain: kringle homology <KRG>  
F;373-407/Product: hepatocyte growth factor activator light chain #status experimental  
F;408-655/Product: hepatocyte growth factor activator heavy chain #status experimental  
F;408-641/Domain: trypsin homology <TRY>  
F;40,48,290,468,492,546/Binding site: carbohydrate (Asn) (covalent) #status predicted  
F;164-155,169-186,188-197,202-330,228-237,245-256,250-267,269-278,286-367,307-349,338-368,447,497,598/Active site: His, Asp, Ser #status predicted

Query Match 39.5%; Score 141; DB 1; Length 655;  
Best Local Similarity 45.6%; Pred. No. 3.3e-08;  
Matches 31; Conservative 6; Mismatches 21; Indels 10; Gaps 3;

QY 1 CFWNQHLHYREDQTSAPAGRLUWLDAQGLASAPVS-----LDAQGLASAPVSGAGNHSCRNPD 52  
DB 286 CFLNGTGYRGVASTASGLSLANSDLLYQELHVS-VGAAALLGLGPAYCRNPDND 344  
QY 53 PRGPCWCV 60  
DB 345 ER-PWCYV 351

C:Superfamily: tissue plasminogen activator; EGF homology; fibronectin type I repeat hom  
C:Keywords: fibrinolysis; glycoprotein; hydrolase; kringle; serine proteinase  
F:18-29/Domain: signal sequence #status predicted <SIG>  
F:30-559/Product: t-plasminogen activator #status predicted <WAT>  
F:30-308/Product: t-plasminogen activator chain A #status predicted <ACH>  
F:38-75/Domain: fibronectin type I repeat homology <IFA>  
F:124-205/Domain: kringle homology <KRI>  
F:213-294/Domain: kringle homology <KR2>  
F:309-559/Product: t-plasminogen activator chain B #status predicted <BCH>  
F:309-553/Domain: trypsin homology <TRY>  
F:38-68-66-75-83-94-88-105-107-116-124-205-145-187-176-200-213-294-234-276-265-289-297-4  
F:149-181/Binding site: carbohydrates (Asn) (covalent) #status predicted  
F:308-309/Cleavage site: Arg-ile (plasmin, trypsin) #status predicted  
F:355-404/510/Active site: His, Asp, Ser #status predicted

Query Match 38.1%; Score 136; DB 1; Length 559;  
Best Local Similarity 40.6%; Pred. No. 1e-07;  
Matches 28; Conservative 8; Mismatches 21; Indels 12; Gaps 3;

QY 1 CFWDNGHLYREDQTSPPAPGLRCLNWLDAQSLGASAPVS-----GAGNHSGYCRNPDE 51  
DB 124 CPEEQGYTYRGTWSTSGAECINW--NSSVLSPYNARRRNPATKLGHNHNYCRNPDR 181

QY 52 DPRGPWCYV 60  
DB 182 DLK-PWCYV 189

RESULT 6  
JS0599  
t-plasminogen activator (EC 3.4.21.68) beta precursor - common vampire bat  
N:Alternate names: tissue plasminogen activator  
C:Species: Desmodus rotundus (common vampire bat)  
C:Date: 31-Mar-1992 #sequence\_revision 31-Mar-1992 #text\_change 16-Jul-1999  
C:Accession: JS0599  
R:Kraetzschmar, J.; Haendler, B.; Langer, G.; Boidol, W.; Bringmann, P.; Alagon, A.; Don  
Gene 105, 229-237, 1991  
A:Title: The plasminogen activator family from the salivary gland of the vampire bat Des  
A:Reference number: JS0597; MUID:92039036; PMID:1937019  
A:Accession: JS0599  
A:Molecule type: mRNA  
A:Residues: 1-431 <KRA>  
C:Cross-references: GB:M63989; NID:G166076; PIDN:AAA1594.1; PID:G166077  
C:Superfamily: tissue plasminogen activator; EGF homology; fibronectin type I repeat hom  
C:Keywords: fibrinolysis; glycoprotein; hydrolase; kringle; serine proteinase  
F:1-21/Domain: signal sequence #status predicted <SIG>  
F:22-36/Domain: propeptide #status predicted <PRO>  
F:37-431/Product: plasminogen activator beta #status predicted <PLA>  
F:41-74/Domain: EGF homology <EGF>  
F:82-163/Domain: kringle homology <KRG>  
F:180-425/Domain: trypsin homology <TRY>  
F:41-52-46-63-65-74-82-163-103-134-158-168-299-211-227-219-288-313-388/Disulfide bon  
F:139-352/Binding site: carbohydrate (Asn) (covalent) #status predicted  
F:179-180/Cleavage site: His-Ser (plasmin) #status predicted  
F:226-275-382/Active site: His, Asp, Ser #status predicted  
F:345-361-378-406/Disulfide bonds: #status predicted

Query Match 37.5%; Score 134; DB 2; Length 431;  
Best Local Similarity 38.6%; Pred. No. 1.3e-07;  
Matches 27; Conservative 9; Mismatches 20; Indels 14; Gaps 3;

QY 1 CFWDNGHLYREDQTSPPAPGLRCLNWLDAQSLG-----ASAPVSGAGNHSGYCRNP 50  
DB 82 CYKQGYTYRGTWSTSGAECINW---NSNLLTRTYNGRRSDATLGLGNHNYCRNP 138

QY 51 EDRGPWCYV 60  
DB 139 NNSK-PWCYV 147

RESULT 7

A34369  
t-plasminogen activator (EC 3.4.21.68) precursor - false vampire bat (Megaderma lyra)  
C:Species: Megaderma lyra  
C:Date: 10-Sep-1999 #sequence\_revision 10-Sep-1999 #text\_change 10-Sep-1999  
C:Accession: A34369  
R:Gardell, S.J.; Duong, L.T.; Diehl, R.E.; York, J.D.; Hare, T.R.; Register, R.B.; Jaco  
J. Biol. Chem. 264, 17947-17952, 1989  
A:Title: Isolation, characterization, and cDNA cloning of a vampire bat salivary plasmi  
A:Reference number: A34369; MUID:90036867; PMID:2509450  
A:Accession: A34369  
A:Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-477 <GAR>  
C:Cross-references: GB:J05082; NID:G166080; PIDN:AAA1596.1; PID:G166081  
C:Superfamily: tissue plasminogen activator; EGF homology; fibronectin type I repeat ho  
C:Keywords: fibrinolysis; glycoprotein; hydrolase; kringle; serine proteinase  
F:1-21/Domain: signal sequence #status predicted <SIG>  
F:22-36/Domain: propeptide #status predicted <PRO>  
F:37-477/Product: plasminogen activator #status predicted <PLA>  
F:42-79/Domain: fibronectin type I repeat homology <IFA>  
F:87-120/Domain: EGF homology <EGF>  
F:128-209/Domain: kringle homology <KRG>  
F:226-471/Domain: trypsin homology <TRY>  
F:442-72/70-79-87-98-92-109-111-120-128-209-149-191-180-204-214-345-257-273-265-334-359-  
F:272-321-428/Active site: His, Asp, Ser #status predicted

Query Match 37.5%; Score 134; DB 1; Length 477;  
Best Local Similarity 38.6%; Pred. No. 1.5e-07;  
Matches 27; Conservative 9; Mismatches 20; Indels 14; Gaps 3;

QY 1 CFWDNGHLYREDQTSPPAPGLRCLNWLDAQSLG-----ASAPVSGAGNHSGYCRNP 50  
DB 128 CYKQGYTYRGTWSTSGAECINW---NSNLLTRTYNGRRSDATLGLGNHNYCRNP 184

QY 51 EDRGPWCYV 60  
DB 185 NNSK-PWCYV 193

RESULT 8  
JS0598  
t-plasminogen activator (EC 3.4.21.68) alpha-2 precursor - common vampire bat  
N:Alternate names: tissue plasminogen activator  
C:Species: Desmodus rotundus (common vampire bat)  
C:Date: 31-Mar-1992 #sequence\_revision 31-Mar-1992 #text\_change 16-Jul-1999  
C:Accession: JS0598  
R:Kraetzschmar, J.; Haendler, B.; Langer, G.; Boidol, W.; Bringmann, P.; Alagon, A.; Do  
Gene 105, 229-237, 1991  
A:Title: The plasminogen activator family from the salivary gland of the vampire bat De  
A:Reference number: JS0597; MUID:92039036; PMID:1937019  
A:Accession: JS0598  
A:Molecule type: mRNA  
A:Residues: 1-477 <KRA>  
C:Cross-references: GB:M63988; NID:G166074; PIDN:AAA1593.1; PID:G166075  
C:Superfamily: tissue plasminogen activator; EGF homology; fibronectin type I repeat ho  
C:Keywords: fibrinolysis; glycoprotein; hydrolase; kringle; serine proteinase  
F:1-21/Domain: signal sequence #status predicted <SIG>  
F:22-36/Domain: propeptide #status predicted <PRO>  
F:37-477/Product: plasminogen activator alpha-2 #status predicted <PLA>  
F:42-79/Domain: fibronectin type I repeat homology <IFA>  
F:87-120/Domain: EGF homology <EGF>  
F:128-209/Domain: kringle homology <KRG>  
F:226-471/Domain: trypsin homology <TRY>  
F:442-72-70-79-87-98-92-109-111-120-128-209-149-191-180-204-214-345-257-273-265-334-359-  
F:185-398/Binding site: carbohydrate (Asn) (covalent) #status predicted  
F:225-226/Cleavage site: His-Ser (plasmin) #status predicted  
F:272-321-428/Active site: His, Asp, Ser #status predicted

Query Match 37.5%; Score 134; DB 2; Length 477;  
Best Local Similarity 38.6%; Pred. No. 1.5e-07;  
Matches 27; Conservative 9; Mismatches 20; Indels 14; Gaps 3;

QY 1 CFWDNGHLYREDQTSPPAPGLRCLNWLDAQSLG-----ASAPVSGAGNHSGYCRNP 50

```

Db 128 CYKQGVTVRGWTSFSSGAQCINW--NSNLLTRRTYNGRRSDAITLGLGHNVCNRPD 184
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 51 EDPGPWCYCV 60
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 185 NNSK-PWCYV 193

RESULT 9
URXAY
u-plasminogen activator (EC 3.4.21.73) precursor - yellow baboon
C:Species: Papio cynocephalus, Papio hamadryas cynocephalus (yellow baboon)
C>Date: 31-Dec-1991 #sequence_revision 31-Dec-1991 #text_change 18-Jun-1999
C:Accession: S14687; S08651
R:Au, Y.P.T.; Wang, T.W.; Clowes, A.W.
Nucleic Acids Res. 18, 3411, 1990
A:Title: Nucleotide and deduced amino acid sequences of baboon urokinase-type plasminogen
A:Reference number: S14687; MUID:90287734; PMID:2113276
A:Accession: S14687
A:Molecule type: mRNA
A:Residues: 1-433 <AU>
A:Cross-references: EMBL:X51935; NID:g938130; PIDN:CAA36200.1; PID:g938131
C:Superfamily: urokinase-type plasminogen activator; EGF homology; kringle homology; try
C:Keywords: glycoprotein; heterodimer; hydrolase; kringle; serine proteinase
F:1-20/Domain: signal sequence #status predicted <SIG>
F:21-176/Product: plasminogen activator chain A #status predicted <ACH>
F:30-61/Domain: EGF homology <EGF>
F:69-150/Domain: kringle homology <KRG>
F:178-433/Product: plasminogen activator chain B #status predicted <BCH>
F:167-298,208-224,216-287,315-384,347-363,374-402/Disulfide bonds: #status predicted
F:223,274,378/Active site: His, Asp, Ser #status predicted
F:324/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 37.0%; Score 132; DB 1; Length 433;
Best Local Similarity 40.3%; Pred. No. 2.3e-07;
Matches 27; Conservative 7; Mismatches 25; Indels 8; Gaps 2;

QY 1 CFWDNGLHYREDOTSPAPGLRCLNWLDA-----QSGLASAPVSGAGNHSYCRNPDEDP 53
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 69 CYGNGHFYRGKASTDTMGRSCLAWNSATVLTQTYAHRSDALQLGLGHNVCNRPD-NR 127
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 54 RGPWCYV 60
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 128 RRPWCYV 134

RESULT 10
URXU
u-plasminogen activator (EC 3.4.21.73) precursor [validated] - human
N:Alternate names: cellular plasminogen activator; urokinase; urokinase-type plasminogen
N:Contains: urokinase-type plasminogen activator chain A; urokinase-type plasminogen act
in form
C:Species: Homo sapiens (man)
C>Date: 17-Dec-1982 #sequence_revision 04-Dec-1986 #text_change 15-Sep-2000
C:Accession: A00931; I52209; J70102; A37561; I38102; S65783; A37562; A37563; A37564; A39
R:Riccio, A.; Grimaldi, G.; Verde, P.; Sebastio, G.; Boast, S.; Biasi, F.
Nucleic Acids Res. 13, 2759-2771, 1985
A:Title: The human urokinase-plasminogen activator gene and its promoter.
A:Reference number: A00931; MUID:85215647; PMID:2987867
A:Accession: A00931
A:Molecule type: DNA
A:Residues: 1-431 <RCIC>
A:Cross-references: GB:X02419; NID:g937601; PIDN:CAA26268.1; PID:gl834524
A:Note: the authors translated the codon ATG for residue 214 as Ile
R:Nagamine, Y.; Pearson, D.; Gratkan, M.
Biochem. Biophys. Res. Commun. 132, 563-569, 1985
A:Title: Exon-intron boundary sliding in the generation of two mRNAs coding for porcine
A:Reference number: I52209; MUID:86050639; PMID:3933505
A:Accession: I52209
A:Status: preliminary; translated from GB/EMBL/DBD
A:Molecule type: DNA
A:Residues: 145-161 <NAG1>

```

```

A:Cross-references: GB:K03027; NID:g340174; PIDN:AAA61257.1; PID:g340175
R:Nagai, M.; Hiramatsu, R.; Kaneda, T.; Hayasuke, N.; Arimura, H.; Nishida, M.; Suyama,
Gene 36, 183-188, 1985
A:Title: Molecular cloning of cDNA coding for human preprourokinase.
A:Reference number: J70102; MUID:86056954; PMID:2415429
A:Accession: J70102
A:Molecule type: mRNA
A:Residues: 1-213, 'I', 215-431 <NAG2>
A:Cross-references: GB:K03226; NID:g340155; PIDN:AA097138.1; PID:g340158; GB:D00244; NI
R:Verde, P.; Stoppelli, M.P.; Galeffi, P.; Di Nocera, P.; Biasi, F.
Proc. Natl. Acad. Sci. U.S.A. 81, 4727-4731, 1984
A:Title: Identification and primary sequence of an unspliced human urokinase poly(A)+ F
A:Reference number: A37561; MUID:84272706; PMID:6589620
A:Accession: A37561
A:Molecule type: mRNA
A:Residues: 66-431 <VER>
A:Cross-references: GB:D00244; NID:g220138
R:Jacobs, P.; Cravador, A.; Loriau, R.; Brockly, F.; Colau, B.; Chuchana, P.; van Elser
DNA 4, 139-146, 1985
A:Title: Molecular cloning, sequencing, and expression in Escherichia coli of human pre
A:Reference number: I38102; MUID:85203359; PMID:3888571
A:Accession: I38102
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-150, 'W', 152-213, 'I', 215-385, 'C', 387-429, 'V', 431 <JAC>
A:Cross-references: EMBL:X02760; NID:g35297; PIDN:CAA26535.1; PID:g35298
R:Yoshimoto, M.; Ushiyama, Y.; Sakai, M.; Tamaki, S.; Hara, H.; Takahashi, K.; Sawasaki
Biochim. Biophys. Acta 1293, 83-89, 1996
A:Title: Characterization of single chain urokinase-type plasminogen activator with a r
A:Reference number: S65783; MUID:96186279; PMID:8652631
A:Accession: S65783
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 21-140, 'L', 142-213, 'I', 215-431 <YOS>
A:Cross-references: EMBL:D11143; NID:gl311467; PIDN:BAA01919.1; PID:gl199928
R:Gunzler, W.A.; Steffens, G.J.; Ottling, F.; Km, S.M.A.; Frankus, E.; Flohe, L.
Hoppe-Seyler's Z. Physiol. Chem. 363, 1155-1165, 1982
A:Title: The primary structure of high molecular mass urokinase from human urine.
A:Reference number: A37562; MUID:83055084; PMID:6754569
A:Accession: A37562
A:Molecule type: protein
A:Residues: 21-177 <GUN>
R:Schaller, J.; Nick, H.; Rickli, E.E.; Gillesen, D.; Lergier, W.; Studer, R.O.
Eur. J. Biochem. 125, 251-257, 1982
A:Title: Human low-molecular-weight urinary urokinase. Partial characterization and pre
A:Reference number: A37563; MUID:83003608; PMID:6749491
A:Accession: A37563
A:Molecule type: protein
A:Residues: 156-176; 179-193, 'T', 195, 'T', 197-224 <SCH>
R:Steffens, G.J.; Gunzler, W.A.; Ottling, F.; Frankus, E.; Flohe, L.
Hoppe-Seyler's Z. Physiol. Chem. 363, 1043-1058, 1982
A:Title: The complete amino acid sequence of low molecular mass urokinase from human ur
A:Reference number: A37564; MUID:83055099; PMID:6754572
A:Accession: A37564
A:Molecule type: protein
A:Residues: 158-410 <STE>
R:Kenzler, E.J.; Buko, A.; Menon, G.; Sarin, V.K.
Biochem. Biophys. Res. Commun. 171, 401-406, 1990
A:Title: Carbohydrate composition and presence of a fucose-protein linkage in recombin
A:Reference number: A35689; MUID:90365737; PMID:2393398
A:Accession: A35689
A:Molecule type: protein
A:Residues: 21-30, 'X', 32, 'X', 34-38, 'X', 40-43 <KEN>
A:Note: identification of a fucose and attempt to determine its attachment site
R:Rabban, S.A.; Desjardins, J.; Bell, A.W.; Banville, D.; Mazar, A.; Henkin, J.; Goltz
Biochem. Biophys. Res. Commun. 173, 1058-1064, 1990
A:Title: An amino-terminal fragment of urokinase isolated from a prostate cancer cell 1
A:Reference number: A36697; MUID:91097529; PMID:2125213
A:Accession: A36697
A:Molecule type: protein
A:Residues: 21-34 <RAS>
R:Li, X.; Bokman, A.M.; Llinas, M.; Smith, R.A.G.; Dobson, C.M.
submitted to the Brookhaven Protein Data Bank, July 1993

```

A;Reference number: A51255; PDB:1KDUU  
 A;Contents: annotation; conformation and disulfide bond assignments by (1)H-NMR, residue  
 R.Li, X.; Smith, R.A.G.; Dobson, C.M.  
 Biochemistry 31, 9562-9571, 1992  
 A;Title: Sequential (1)H NMR assignments and secondary structure of the kringle domain of  
 u-plasminogen activator; A44375; PMID:93003110; PMID:1327118  
 A;Contents: annotation; conformation and disulfide bond assignments by (1)H-NMR  
 R.Hansen, A.P.; Petros, A.M.; Meadows, R.P.; Nettesheim, D.G.; Mazur, A.P.; Olejniczak,  
 submitted to the Brookhaven Protein Data Bank, January 1994  
 A;Reference number: A66822; PDB:1URK  
 A;Contents: annotation; conformation and disulfide bond assignments by (1)H-NMR, residue  
 R.Spraggon, G.S.; Phillips, C.; Nowak, U.K.; Ponting, C.P.; Saunders, D.; Dobson, C.M.;  
 submitted to the Brookhaven Protein Data Bank, July 1995  
 A;Reference number: A66058; PDB:1LMW  
 A;Contents: annotation; X-ray crystallography, 2.5 angstroms, residues 168-175; 179-426  
 C;Comment: This enzyme is found in urine in a high molecular mass form, consisting of A  
 C;Comment: Urokinase-type plasminogen activator proteolytically activates plasminogen, a  
 C;Genetics:  
 A;Gene: GDB:PLAU  
 A;Cross-references: GDB:119497; OMIM:191840  
 A;Map position: 10q24-10q24  
 A;Introns: 19/3, 29/1, 65/1, 123/2, 154/1, 227/2, 277/1, 324/1, 373/3  
 C;Function:  
 A;Description: proteolytically activates plasminogen  
 A;Pathway: fibrinolysis  
 C;Superfamily: urokinase-type plasminogen activator; EGF homology; kringle homology; try  
 C;Keywords: fibrinolysis; glycoprotein; heterodimer; hydrolase; kringle; serine proteina  
 F;1-20/Domain: signal sequence #status predicted <SIG>  
 F;21-431/Product: urokinase-type plasminogen activator, single chain form #status predic  
 F;21-177/Product: urokinase-type plasminogen activator chain A #status experimental <MPA  
 F;31-62/Domain: EGF homology <EGF>  
 F;151/Domain: kringle homology <KR>  
 F;156-177/Product: urokinase-type plasminogen activator chain A1 #status experimental <M  
 F;179-431/Product: urokinase-type plasminogen activator chain B #status experimental <M  
 F;31-39, 33-51, 53-62, 70-151, 91-133, 122-146, 168-299, 209-225, 217-288, 313-382, 345-361, 372-40  
 F;38/Binding site: carboxydrate (Thr) (covalent) #status predicted  
 F;178-179/Cleavage site: Lys-Ile (plasmin) #status experimental  
 F;224,275,376/Active site: His, Asp, Ser #status experimental  
 F;322/Binding site: carboxydrate (Asn) (covalent) #status experimental

Query Match 36.7%; Score 131; DB 1; Length 431;  
 Best Local Similarity 40.3%; Pred. No. 3e-07;  
 Matches 27; Conservative 7; Mismatches 25; Indels 8; Gaps 2;  
 QY 1 CFWDNGHLYREDQTSAPGLRCLNWDA-----OSGLASAPVSGAGNHSYCRNPDEDP 53  
 Db 70 CYEGNGHFYRGKASTDTCGRPCLPWNSATVLOQTYHAHRSALQLGLGKHNYCRNPD-NR 128

QY 54 RGPWCYV 60  
 Db 129 RRPWCYV 135

RESULT 11  
 UKPG  
 u-plasminogen activator (EC 3.4.21.73) precursor - pig  
 N;Alternate names: uPA  
 C;Species: Sus scrofa domestica (domestic pig)  
 C;Date: 04-Dec-1986 #sequence\_revision 17-Mar-1987 #text\_change 07-Aug-1998  
 C;Accession: A00932  
 R;Nagamine, Y.; Pearson, D.; Altus, M.S.; Reich, E.  
 Nucleic Acids Res. 12, 9525-9541, 1984  
 A;Title: cDNA and gene nucleotide sequence of porcine plasminogen activator.  
 A;Reference number: A00932; MUID:95087954; PMID:6096832  
 A;Accession: A00932  
 A;Molecule type: DNA  
 A;Residues: 1-240; 'H', 242-442 <NAG1>  
 A;Experimental source: kidney cell line LLC-PK1  
 R;Nagamine, Y.  
 submitted to the Protein Sequence Database, December 1986  
 A;Reference number: A37566  
 A;Contents: annotation; correction to residue 241

C;Genetics:  
 A;Introns: 19/3, 31/1, 67/1, 125/2, 165/1, 238/2, 288/1, 335/1, 384/3  
 C;Superfamily: urokinase-type plasminogen activator; EGF homology; kringle homology; tr  
 C;Keywords: glycoprotein; heterodimer; hydrolase; kringle; serine proteinase  
 F;1-20/Domain: signal sequence #status predicted <SIG>  
 F;21-188/Product: urokinase-type plasminogen activator chain A #status predicted <ACH>  
 F;33-64/Domain: EGF homology <EGF>  
 F;72-153/Domain: kringle homology <KR>  
 F;150-442/Product: urokinase-type plasminogen activator chain B #status predicted <BCH>  
 F;150-430/Domain: trypsin homology <TRY>  
 F;152/Binding site: carboxydrate (Asn) (covalent) #status predicted  
 F;179-310, 220-236, 228-299, 324-393, 356-372, 383-411/Disulfide bonds: #status predicted  
 F;235, 286, 387/Active site: His, Asp, Ser #status predicted

Query Match 36.4%; Score 130; DB 1; Length 442;  
 Best Local Similarity 44.8%; Pred. No. 3.9e-07;  
 Matches 30; Conservative 4; Mismatches 25; Indels 8; Gaps 3;

QY 1 CFWDNGHLYREDQTSAPGLRCLNWDAOSGL-----ASAPVS---CAGNHSYCRNPDEDP 53  
 Db 72 CFEGNGHSYRGKANTTGRPCLPWNSATVLLTYHAHRPDALQLGLGKHNYCRNPD-NQ 130

QY 54 RGPWCYV 60  
 Db 131 RRPWCYV 137

RESULT 12  
 JC5061  
 macrophage-stimulating protein 1 precursor - rat  
 C;Species: Rattus norvegicus (Norway rat)  
 C;Date: 31-Jan-1997 #sequence\_revision 31-Jan-1997 #text\_change 16-Jun-2000  
 C;Accession: JC5061  
 R;Onishi, K.; Iwama, A.; Matsuno, K.; Szaki, T.; Sakamoto, O.; Hamaguchi, I.; Takasu,  
 Biochem. Biophys. Res. Commun. 227, 273-280, 1996  
 A;Title: Molecular cloning of rat macrophage-stimulating protein and its involvement in  
 A;Reference number: JC5061; MUID:97011136; PMID:8858136  
 A;Accession: JC5061  
 A;Molecule type: mRNA  
 A;Residues: 1-716 <OHS>

A;Cross-references: EMBL:X95096; NID:G1669718; PID:CAA64473.1; PID:G1669719  
 C;Complex: disulfide-bonded heterodimer of chains derived from the same precursor  
 C;Superfamily: hepatocyte growth factor; kringle homology; trypsin homology  
 C;Keywords: duplication; glycoprotein; growth factor; kringle  
 F;1-31/Domain: signal sequence #status predicted <SIG>  
 F;32-488, 489-716/Product: macrophage-stimulating protein 1 #status predicted <MAT>  
 F;32-488/Domain: macrophage-stimulating protein 1 alpha chain #status predicted <ACH>  
 F;110-186/Domain: kringle homology <KR1>  
 F;191-268/Domain: kringle homology <KR12>  
 F;292-370/Domain: kringle homology <KR13>  
 F;379-457/Domain: kringle homology <KR14>  
 F;489-716/Domain: macrophage-stimulating protein 1 beta chain #status predicted <BCH>  
 F;489-709/Domain: trypsin homology <TRY>  
 F;72,305,620/Binding site: carboxydrate (Asn) (covalent) #status predicted

Query Match 35.9%; Score 128; DB 1; Length 716;  
 Best Local Similarity 41.9%; Pred. No. 1.1e-06;  
 Matches 26; Conservative 3; Mismatches 29; Indels 4; Gaps 2;

QY 1 CFWDNGHLYREDQTSAPGLRCLNW---LDAQSLASAPVSGAGNHSYCRNPDEDPGFW 57  
 Db 110 CIMONGASYRGTVARTADGLPCQASRRFFNDHKYPTPKNGI-EENFCRNPDPGFW 168

QY 58 CY 59  
 Db 169 CY 170

RESULT 13  
 JN0560  
 u-plasminogen activator (EC 3.4.21.73) precursor - bovine  
 N;Alternate names: uPA  
 C;Species: Bos primigenius taurus (cattle)

C>Date: 31-Dec-1993 #sequence\_revision 31-Dec-1993 #text\_change 16-Jul-1999  
C/Accession: JN0560  
R/Kraetzschmar, J.; Haendler, B.; Kojima, S.; Rifkin, D.B.; Schleuning, W.D.  
Gene 125, 177-183, 1993  
A/Title: Bovine urokinase-type plasminogen activator and its receptor: cloning and induction  
A/Reference number: JN0560; MUID:932161119; PMID:8385052  
A/Accession: JN0560  
A/Molecule type: mRNA  
A/Residues: 1-433 <KRA>  
A/Cross-references: GB:I03546; NID:G163800; PIDN:AAA51419.1; PID:G163801  
C/Superfamily: urokinase-type plasminogen activator; EGF homology; kringle homology; try  
C/Keywords: glycoprotein; heterodimer; hydrolase; kringle; serine proteinase  
F:1-20/Domain: signal sequence #status predicted <SIG>  
F:21-179/Product: plasminogen activator chain A #status predicted <MA1>  
F:21-179/Product: urokinase-type plasminogen activator chain A #status predicted <ACH>  
F:33-64/Domain: EGF homology <EGF>  
F:72-153/Domain: kringle homology <KRG>  
F:181-433/Product: plasminogen activator chain B #status predicted <MA2>  
F:181-421/Domain: trypsin homology <TRY>  
F:170-301,211-227,219-290,315-384,347-363,374-402/Disulfide bonds: #status predicted  
F:226,277,378/Active site: His, Asp, Ser #status predicted  
Query Match 35.6%; Score 127; DB 1; Length 433;  
Best Local Similarity 41.2%; Pred. No. 8.5e-07;  
Matches 28; Conservative 6; Mismatches 24; Indels 10; Gaps 3;  
QY 1 CFWDNGHLYREDQTSAPGLRCLNWLDAQSGL-----ASAPVSGAGNHSYCRNPDED 52  
DB 72 CYQNGHSYKYGKANKRDLGSRPCLAW-DSPTVLLKMYHRSDAIQGLGKHYCRNPD-N 129  
QY 53 PRGPWCYV 60  
DB 130 QRPPWCYV 137  
RESULT 14  
JN0600  
t-plasminogen activator (EC 3.4.21.68) gamma precursor - common vampire bat  
N/Alternate names: tissue plasminogen activator  
C/Species: Desmodus rotundus (common vampire bat)  
C/Date: 31-Mar-1992 #sequence\_revision 31-Mar-1992 #text\_change 16-Jul-1999  
C/Accession: JN0600  
R/Kraetzschmar, J.; Haendler, B.; Langer, G.; Boidol, W.; Bringmann, P.; Alagon, A.; Don  
Gene 105, 229-237, 1991  
A/Title: The plasminogen activator family from the salivary gland of the vampire bat Des  
A/Reference number: JN0597; MUID:92039036; PMID:1937019  
A/Accession: JN0600  
A/Molecule type: mRNA  
A/Residues: 1-394 <KRA>  
A/Cross-references: GB:M63990; NID:G166078; PIDN:AAA1595.1; PID:G166079  
A/Note: the authors translated the codon ATC for residue 75 as Thr  
C/Superfamily: tissue plasminogen activator; EGF homology; fibronectin type I repeat hom  
C/Keywords: fibrinolysis; glycoprotein; hydrolase; kringle; serine proteinase  
F:1-21/Domain: signal sequence #status predicted <SIG>  
F:22-36/Domain: propeptide #status predicted <PRO>  
F:37-394/Product: propeptide #status predicted <PLA>  
F:45-126/Domain: kringle homology <KRG>  
F:143-388/Domain: trypsin homology <TRY>  
F:45-126,66-108,97-121,131-262,174-190,182-251,276-351,308-324,341-369/Disulfide bonds:  
F:142-143/Cleavage site: His-Ser (plasmin) #status predicted  
F:189,238,345/Active site: His, Asp, Ser #status predicted  
F:315/Binding site: carbohydrate (Asn) (covalent) #status predicted  
Query Match 35.3%; Score 126; DB 2; Length 394;  
Best Local Similarity 37.3%; Pred. No. 1e-06;  
Matches 25; Conservative 8; Mismatches 26; Indels 8; Gaps 2;  
QY 1 CFWDNGHLYREDQTSAPGLRCLNW-----LDAQSGLASAPVSGAGNHSYCRNPDED 53  
DB 45 CYKDGVTYGTWSTSGAQCNWNSNLLIRTYNGRMPKAVKLGKGNHYCRNPDGAS 104  
QY 54 RGPWCYV 60  
:

DB 105 K-PWCYV 110  
RESULT 15  
A40332  
macrophage-stimulating protein 1 precursor - mouse  
N/Alternate names: hepatocyte growth factor-like protein  
C/Species: Mus musculus (house mouse)  
C/Date: 17-Jul-1992 #sequence\_revision 17-Jul-1992 #text\_change 18-Jun-1999  
C/Accession: A40332; B40332  
R/Degen, S.J.F.; Stuart, L.A.; Han, S.; Jamison, C.S.  
Biochemistry 30, 9781-9791, 1991  
A/Title: Characterization of the mouse cDNA and gene coding for a hepatocyte growth fac  
A/Reference number: A40332; MUID:9202017; PMID:1832957  
A/Accession: A40332  
A/Molecule type: DNA  
A/Residues: 1-716 <DEG>  
A/Cross-references: GB:M74180; NID:G193831; PIDN:AAA50166.1; PID:G193832  
A/Accession: B40332  
A/Molecule type: mRNA  
A/Residues: 1-18, P', 20-716 <DEG2>  
A/Cross-references: GB:M74181; NID:G193833; PIDN:AAA50167.1; PID:G193834  
C/Genetics:  
A/Intons: 18/1; 67/2; 105/1; 143/2; 189/1; 229/2; 269/1; 334/2; 378/1; 412/2; 458/1;  
C/Complex: disulfide-bonded heterodimer of chains derived from the same precursor  
C/Superfamily: hepatocyte growth factor; kringle homology; trypsin homology  
C/Keywords: duplication; glycoprotein; growth factor; kringle  
F:1-31/Domain: signal sequence #status predicted <SIG>  
F:19-488,489-716/Product: macrophage-stimulating protein 1 #status experimental <MAT>  
F:110-186/Domain: alpha chain #status experimental <ACH>  
F:191-268/Domain: kringle homology <KR1>  
F:292-370/Domain: kringle homology <KR2>  
F:379-457/Domain: kringle homology <KR4>  
F:484-711/Domain: beta chain #status experimental <BCH>  
F:489-709/Domain: trypsin homology <TRY>  
F:72,173,305,620/Binding site: carbohydrate (Asn) (covalent) #status predicted  
Query Match 35.3%; Score 126; DB 1; Length 716;  
Best Local Similarity 41.9%; Pred. No. 1.8e-06;  
Matches 26; Conservative 3; Mismatches 29; Indels 4; Gaps 2;  
QY 1 CFWDNGHLYREDQTSAPGLRCLNW-----LDAQSGLASAPVSGAGNHSYCRNPDED 57  
DB 110 CIMDNGSVYRGTVARTAGGLPCQAWSRFPNDHKYTPKNGL-EENFCRNPDPGRGPW 168  
QY 58 CY 59  
DB 169 CY 170  
Search completed: March 17, 2004, 07:07:59  
Job time : 5.19919 secs

GenCore version 5.1.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 17, 2004, 06:55:43 ; Search time 2.42261 Seconds  
(without alignments)  
1289.604 Million cell updates/sec

Title: US-10-057-951-2\_COPY\_25\_84  
Perfect score: 357  
Sequence: 1 CFWDNHLYREDQTSFAPGL.....GNHSYCRNPDEDPGRPCVY 60

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SwissProt\_42:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	154	43.1	566	TPA_BOVIN	Q28198 bos taurus
2	146	40.9	562	TPA_HUMAN	P00750 homo sapien
3	146	40.9	653	HGFL_MOUSE	Q90988 mus musculus
4	141	39.5	559	TPA_RAT	P19637 rattus norv
5	141	39.5	655	HGFL_HUMAN	Q04756 homo sapien
6	136	38.1	559	TPA_MOUSE	P11214 mus musculus
7	134	37.5	431	URTB_DESRO	P98121 desmodus ro
8	134	37.5	477	URT2_DESRO	P15638 desmodus ro
9	132	37.0	433	UROK_PAPCY	P16227 papio cynoc
10	131	36.7	431	UROK_HUMAN	P00749 homo sapien
11	130	36.4	442	UROK_PIG	P04185 sus scrofa
12	127	35.6	433	UROK_BOVIN	Q05589 bos taurus
13	126	35.3	394	URTG_DESRO	P49150 desmodus ro
14	126	35.3	716	HGFL_MOUSE	P26928 mus musculus
15	123	34.5	169	PLMN_RAT	Q01177 rattus norv
16	121	33.9	810	PLMN_HUMAN	P00747 homo sapien
17	119	33.3	433	UROK_MOUSE	P06869 mus musculus
18	118	33.3	728	HGF_HUMAN	P14210 homo sapien
19	118	33.1	432	UROK_RAT	P29598 rattus norv
20	118	33.1	728	HGF_MOUSE	Q08048 mus musculus
21	118	33.1	810	PLMN_ERIEU	Q29485 erinaceus e
22	117.5	32.9	461	KRM2_MOUSE	Q8k1s7 mus musculus
23	117.5	32.9	462	KRM2_HUMAN	Q8ncw0 homo sapien
24	115	32.2	477	URT1_DESRO	P98119 desmodus ro
25	113.5	31.8	333	PLMN_CANFA	P80009 canis famil
26	113.5	31.8	790	PLMN_PIG	P06867 sus scrofa
27	113.5	31.8	4548	APOA_HUMAN	P08519 homo sapien
28	113	31.7	452	KRM1_XENLA	Q90y90 xenopus lae
29	113	31.7	810	PLMN_MACMU	P12545 macaca mula
30	112	31.4	728	HGF_RAT	P17945 rattus norv
31	111.5	31.2	603	FAI2_CAVPO	Q04962 cavia porce
32	111	31.1	812	PLMN_BOVIN	P06868 bos taurus
33	110	30.8	473	KRM1_MOUSE	Q99n43 mus musculus

ALIGNMENTS

RESULT 1

ID	TPA_BOVIN	STANDARD;	PRT;	566 AA.
AC	Q28198;			
DT	01-NOV-1997 (Rel. 35, Created)			
DT	01-NOV-1997 (Rel. 35, Last sequence update)			
DT	15-MAR-2004 (Rel. 43, Last annotation update)			
DE	Tissue-type plasminogen activator precursor (EC 3.4.21.68) (tPA)			
DE	(t-PA) (t-plasminogen activator).			
GN	PLAT.			
OS	Bos taurus (Bovine).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;			
OC	Bovidae; Bovinae; Bos.			
OX	NCBI_TaxID=9913;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	TISSUE=Kidney;			
RA	Ravn P., Berglund L., Petersen T.E.;			
RT	"Cloning and characterization of the bovine plasminogen activators uPA			
RT	and tPA.";			
RL	Int. Dairy J. 5:605-617(1995).			
CC	-!- FUNCTION: Converts the abundant, but inactive, zymogen plasminogen			
CC	to plasmin by hydrolyzing a single Arg-Val bond in plasminogen. By			
CC	controlling plasmin-mediated proteolysis, it plays an important			
CC	role in tissue remodeling and degradation, in cell migration and			
CC	many other physiological events.			
CC	-!- CATALYTIC ACTIVITY: Specific cleavage of Arg- -Val bond in			
CC	plasminogen to form plasmin.			
CC	-!- SUBUNIT: Heterodimer of chain A and chain B held by a disulfide			
CC	bond.			
CC	-!- SUBCELLULAR LOCATION: Secreted; extracellular.			
CC	-!- PTM: THE SINGLE CHAIN, ALMOST FULLY ACTIVE ENZYME, CAN BE FURTHER			
CC	PROCESSED INTO A TWO-CHAIN FULLY ACTIVE FORM BY A CLEAVAGE AFTER			
CC	ARG-314 CATALYZED BY PLASMIN, TISSUE KALLIKREIN OR FACTOR XA.			
CC	-!- MISCELLANEOUS: Binds to the kringle structure of the fibrin A			
CC	chain. Binding to fibrin enhances its catalytic activity.			
CC	-!- SIMILARITY: Belongs to peptidase family S1.			
CC	-!- SIMILARITY: Contains 1 EGF-like domain.			
CC	-!- SIMILARITY: Contains 1 fibronectin type I domain.			
CC	-!- SIMILARITY: Contains 2 kringle domains.			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (See <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a>			
CC	or send an email to <a href="mailto:license@isb-sib.ch">license@isb-sib.ch</a> ).			
CC	-----			
CC	EMBL; X85800; CAA59795.1; -			
DR	HSSP; P00750; 1RTF.			
DR	MEROPS; S01.232; -			
DR	InterPro; IPR009003; Cys Ser trypsin.			
DR	InterPro; IPR006209; EGF-like			
DR	InterPro; IPR000083; Fibrinctn.			



```
DR InterPro: IPR006210; IEGF.
DR InterPro: IPR000001; Kringle.
DR InterPro: IPR001254; Peptidase_S1.
DR InterPro: IPR001314; Peptidase_S1A.
DR Pfam: PF00008; EGF_1.
DR Pfam: PF00039; fnl_1.
DR Pfam: PF00051; kringle_2.
DR Pfam: PF00089; trypsin_1.
DR PRINTS: PR00772; CHYMOTRYPSIN.
DR PRINTS: PR00018; KRINGLE.
DR ProDom: PD000395; Kringle_2.
DR SMART: SM00181; EGF_1.
DR SMART: SM00058; FNI_1.
DR SMART: SM00130; KR_2.
DR SMART: SM00020; TRYD_SPC_1.
DR PROSITE: PS00022; EGF_1; 1.
DR PROSITE: PS01186; EGF_2; 1.
DR PROSITE: PS00026; EGF_3; 1.
DR PROSITE: PS01253; FIBRONECTIN_1; 1.
DR PROSITE: PS00021; KRINGLE_1; 1.
DR PROSITE: PS00070; KRINGLE_2; 2.
DR PROSITE: PS0240; TRYPSIN_DOM; 1.
DR PROSITE: PS00134; TRYPSIN_HIS; 1.
DR PROSITE: PS00135; TRYPSIN_SRR; 1.
KW Plasminogen activation; Hydrolase; Serine protease; Glycoprotein;
KW Plasma; Kringle; EGF-like domain; Repeat; Signal.
FT SIGNAL 1 21 BY SIMILARITY.
FT PROPEP 22 33
FT CHAIN 34 566 TISSUE-TYPE PLASMINOGEN ACTIVATOR.
FT CHAIN 34 314 TISSUE-TYPE PLASMINOGEN ACTIVATOR A
FT CHAIN 315 566 TISSUE-TYPE PLASMINOGEN ACTIVATOR B
FT DOMAIN 40 82 FIBRONECTIN TYPE-I.
FT DOMAIN 83 121 EGF-LIKE.
FT DOMAIN 128 209 KRINGLE 1.
FT DOMAIN 219 300 KRINGLE 2.
FT DOMAIN 315 566 SERINE PROTEASE.
FT ACT_SITE 361 361 CHARGE RELAY SYSTEM.
FT ACT_SITE 410 410 CHARGE RELAY SYSTEM.
FT ACT_SITE 517 517 CHARGE RELAY SYSTEM.
FT DISULFID 42 72 BY SIMILARITY.
FT DISULFID 70 79 BY SIMILARITY.
FT DISULFID 87 98 BY SIMILARITY.
FT DISULFID 92 109 BY SIMILARITY.
FT DISULFID 111 120 BY SIMILARITY.
FT DISULFID 128 209 BY SIMILARITY.
FT DISULFID 149 191 BY SIMILARITY.
FT DISULFID 180 204 BY SIMILARITY.
FT DISULFID 219 300 BY SIMILARITY.
FT DISULFID 240 282 BY SIMILARITY.
FT DISULFID 271 295 BY SIMILARITY.
FT DISULFID 303 434 INTERCHAIN (BY SIMILARITY).
FT DISULFID 346 362 BY SIMILARITY.
FT DISULFID 354 423 BY SIMILARITY.
FT DISULFID 448 523 BY SIMILARITY.
FT DISULFID 480 496 BY SIMILARITY.
FT DISULFID 513 541 BY SIMILARITY.
FT CARBOHYD 153 153 N-LINKED (GLCNAC... ) (POTENTIAL).
FT CARBOHYD 487 487 N-LINKED (GLCNAC... ) (POTENTIAL).
SQ SEQUENCE 566 AA; 63701 MW; 2868684E32276C3 CRC64;

Query Match 43.1%; Score 154; DB 1; Length 566;
Best Local Similarity 44.9%; Pred. No. 3.3e-10;
Matches 31; Conservative 6; Mismatches 20; Indels 12; Gaps 3;

Qy 1 CFWNHGLYREDQTSAPGLRCNLNWLDAQSLASAPVS-----GAGNHGYCNPDE 51
Db 128 CYKQGVAYRTWTSAESGAECANW--NSSGLAKPKYGRPRNAILGLGNHYCRNPQ 185
Qy 52 DPGGFWCYV 60
Db 186 DSK-FWCYV 193
```

```
RESULT 2
TPA_HUMAN STANDARD; PRT; 562 AA.
AC P00750; Q15103;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Tissue-type plasminogen activator precursor (EC 3.4.21.68) (tPA)
DE (t-PA) (t-plasminogen activator) (Alteplase) (Retelase).
GN PLAT.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
[1]
RN SEQUENCE FROM N.A.
RP TISSUE=Melanoma;
RC MEDLINE=83115262; PubMed=6337343;
RA Pennica D., Holmes W.E., Kohr W.J., Harkins R.N., Vehar G.A., H.L.,
RA Ward C.A., Bennett W.F., Yelverton E., Seeburg P.H., Heyneker H.L.,
RA Goeddel D.V., Collen D.;
RT "Cloning and expression of human tissue-type plasminogen activator
RT cDNA in E. coli.";
RL Nature 301:214-221(1983).
[2]
RN SEQUENCE FROM N.A.
RP TISSUE=Fetal lung;
RC MEDLINE=88262579; PubMed=3133640;
RA Sasaki H., Saito Y., Hayashi M., Otsuka K., Niwa M.;
RT "Nucleotide sequence of the tissue-type plasminogen activator cDNA
RT from human fetal lung cells.";
RL Nucleic Acids Res. 16:5695-5695(1988).
[3]
RN SEQUENCE FROM N.A.
RP MEDLINE=88054470; PubMed=2824147;
RA Reddy V.B., Garramone A.J., Sasak H., Wei C.-M., Watkins P., Galli J.,
RA Hsiung N.;
RT "Expression of human uterine tissue-type plasminogen activator in
RT mouse cells using BPV vectors.";
RL DNA 6:461-472(1987).
[4]
RN SEQUENCE FROM N.A.
RP MEDLINE=86196143; PubMed=3009482;
RA Frieze Degen S.J., Rajput B., Reich E.;
RT "The human tissue plasminogen activator gene.";
RL J. Biol. Chem. 261:6972-6985(1986).
[5]
RN SEQUENCE FROM N.A.
RP MEDLINE=84298137; PubMed=6089198;
RA Ny T., Elgh F., Lund B.;
RT "The structure of the human tissue-type plasminogen activator gene:
RT correlation of intron and exon structures to functional and
RT structural domains";
RL Proc. Natl. Acad. Sci. U.S.A. 81:5355-5359(1984).
[6]
RN SEQUENCE FROM N.A.
RP MEDLINE=86284200; PubMed=3090401;
RA Harris T.J., Patel T., Marston F.A., Little S., Emtage J.S.,
RA Opdenakker G., Volckaert G., Rombauts W., Billiau A., Somer P.;
RT "Cloning of cDNA coding for human tissue-type plasminogen activator
RT and its expression in Escherichia coli.";
RL Mol. Biol. Med. 3:279-292(1986).
[7]
RN SEQUENCE FROM N.A. (ISOFORM SHORT).
RP TISSUE=Umbilical vein;
RC MEDLINE=90192129; PubMed=2107528;
RA Siebert P.D., Fong K.;
RT "Variant tissue-type plasminogen activator (PLAT) cDNA obtained from
RT human endothelial cells.";
RL Nucleic Acids Res. 18:1086-1086(1990).
[8]
```

RP SEQUENCE FROM N.A.  
RC TISSUE=Brain;  
RX MEDLINE=22388257; PubMed=12477932;  
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.W., Schuler G.D.,  
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,  
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,  
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,  
RA Stapleton M., Soares M.B., Bonaldo M.P., Casavant T.L., Scheetz T.E.,  
RA Brownstein M.J., Ustin T.B., Toshiyuki S., Carninci P., Prange C.,  
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,  
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
RA Richards S., Worley K.C., Hale S.C., Garcia A.M., Gay L.J., Hulyk S.W.,  
RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
RA Fahey J., Helton E., Kettaman M., Madan A., Rodrigues S., Sanchez A.,  
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,  
RA Butterfield Y.S.N., Krzywicki M.I., Skalska U., Smalios D.E.,  
RA Scherch A., Schein J.E., Jones S.J.M., Marra W.A.,  
RT "Generation and initial analysis of more than 15,000 full-length  
RL human and mouse cDNA sequences.";  
RN Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).  
[9]  
RP SEQUENCE OF 212-361 FROM N.A.  
RX MEDLINE=83169656; PubMed=6572897;  
RA Edlund I., Ny T., Raaby M., Heden L.-O., Palm G., Holmgren E.,  
RA Josephson S.;  
RT "Isolation of cDNA sequences coding for a part of human tissue  
RT plasminogen activator.";  
RL Proc. Natl. Acad. Sci. U.S.A. 80:349-352 (1983).  
[10]  
RP SEQUENCE OF 1-36 FROM N.A.  
RX MEDLINE=85289338; PubMed=3161893;  
RA Fisher R., Waller E.K., Grosi G., Thompson D., Tizard R.,  
RA Schleuning W.-D.;  
RT "Isolation and characterization of the human tissue-type plasminogen  
RT activator structural gene including its 5' flanking region.";  
RN J. Biol. Chem. 260:11223-11230 (1985).  
[11]  
RP SEQUENCE OF 31-562 FROM N.A.  
RX MEDLINE=91291340; PubMed=1368681;  
RA Itagaki Y., Yasuda H., Morinaga T., Mitsuda S., Higashio K.;  
RT "Purification and characterization of tissue plasminogen activator  
RT secreted by human embryonic lung diploid fibroblasts, IMR-90 cells.";  
RL Agric. Biol. Chem. 55:1225-1232 (1991).  
[12]  
RP SEQUENCE OF 36-562.  
RC TISSUE=Melanoma;  
RX MEDLINE=85000468; PubMed=6433976;  
RA Pohl G., Kaelstroom M., Bergsdorf N., Wallen P., Joernvall H.;  
RT "Tissue plasminogen activator: peptide analyses confirm an indirectly  
RT derived amino acid sequence, identify the active site serine residue,  
RT establish glycosylation sites, and localize variant differences.";  
RL Biochemistry 23:3701-3707 (1984).  
[13]  
RP SEQUENCE OF 33-52 AND 311-330.  
RC TISSUE=Melanoma;  
RX MEDLINE=83209620; PubMed=6682760;  
RA Wallen P., Pohl G., Bergsdorf N., Raaby M., Ny T., Joernvall H.;  
RT "Purification and characterization of a melanoma cell plasminogen  
RT activator.";  
RL Eur. J. Biochem. 132:681-686 (1983).  
[14]  
RP STRUCTURE OF CARBOHYDRATES.  
RX MEDLINE=90092112; PubMed=2513186;  
RA Pfeiffer G., Schmidt M., Strube K.-H., Geyer R.;  
RT "Carbohydrate structure of recombinant human uterine tissue  
RT plasminogen activator expressed in mouse epithelial cells.";  
RL Eur. J. Biochem. 186:273-286 (1989).  
[15]  
RP CARBOHYDRATE-LINKAGE SITE THR-96.  
RX MEDLINE=91159408; PubMed=1900431;  
RA Harris R.J., Leonard C.K., Guzzetta A.W., Spellman M.W.;  
RT "Tissue plasminogen activator has an O-linked fucose attached to  
RL threonine-61 in the epidermal growth factor domain.";  
RN Biochemistry 30:2311-2314 (1991).  
[16]  
RP DISULFIDE BONDS IN KRINGLE 2  
RX MEDLINE=91244765; PubMed=1645336;  
RA Vlahos C.J., Wilhelm O.G., Hassell T., Jaskunas S.R., Bang N.U.;  
RT "Disulfide pairing of the recombinant kringle-2 domain of tissue  
RL plasminogen activator produced in *Escherichia coli*.";  
RN J. Biol. Chem. 266:10070-10072 (1991).  
[17]  
RP X-RAY CRYSTALLOGRAPHY (2.3 ÅNGSTRÖMS) OF CATALYTIC DOMAIN.  
RX MEDLINE=96200985; PubMed=8613982;  
RA Lamba D., Bauer M., Huber R., Fischer S., Rudolph R., Kohnert U.,  
RA Bode W.;  
RT "The 2.3 Å crystal structure of the catalytic domain of recombinant  
RL two-chain human tissue-type plasminogen activator.";  
RN J. Mol. Biol. 258:117-135 (1996).  
[18]  
RP X-RAY CRYSTALLOGRAPHY (3.1 ÅNGSTRÖMS) OF CATALYTIC DOMAIN.  
RX MEDLINE=97449126; PubMed=9305622;  
RA Renatus M., Engh R.A., Stubbs M.T., Huber R., Fischer S., Kohnert U.,  
RA Bode W.;  
RT "Lysine 156 promotes the anomalous proenzyme activity of tPA: X-ray  
RL crystal structure of single-chain human tPA.";  
RN EMBO J. 16:4797-4805 (1997).  
[19]  
RP X-RAY CRYSTALLOGRAPHY (2.4 ÅNGSTRÖMS) OF KRINGLE 2.  
RX MEDLINE=92118803; PubMed=1310033;  
RA de Vos A., Ultsch M.H., Kelley R.F., Padmanabhan K., Tulinsky A.,  
RA Westbrook M.L., Kosiakof A.A.;  
RT "Crystal structure of the kringle 2 domain of tissue plasminogen  
RL Biochemistry 31:270-279 (1992).  
[20]  
RP STRUCTURE BY NMR OF KRINGLE 2.  
RX MEDLINE=9012799; PubMed=2558718;  
RA Byeon I.-J.L., Kelley R.F., Ilinas M.;  
RT "1H NMR structural characterization of a recombinant kringle 2 domain  
RL from human tissue-type plasminogen activator.";  
RN Biochemistry 28:9350-9360 (1989).  
[21]  
RP STRUCTURE BY NMR OF KRINGLE 2.  
RX MEDLINE=91200042; PubMed=1901789;  
RA Byeon I.-J.L., Kelley R.F., Ilinas M.;  
RT "Kringle-2 domain of the tissue-type plasminogen activator. 1H-NMR  
RL assignments and secondary structure.";  
RN Eur. J. Biochem. 197:155-165 (1991).  
[22]  
RP STRUCTURE BY NMR OF KRINGLE 2.  
RX MEDLINE=92106329; PubMed=1762144;  
RA Byeon I.-J.L., Ilinas M.;  
RT "Solution structure of the tissue-type plasminogen activator kringle  
RL 2 domain complexed to 6-aminohexanoic acid an antifibrinolytic  
RT drug.";  
RN J. Mol. Biol. 222:1035-1051 (1991).  
[23]  
Query Match 40.9%; Score 146; DB 1; Length 562;  
Best Local Similarity 43.5%; Pred. No. 2,7e-09;  
Matches 30; Conservative 5; Mismatches 22; Indels 12; Gaps 3;  
QY 1 CFWDNGLHYREDQTSFAPGLRCLNWLDAQSLASAPVS-----GAGNHSYCRNPDE 51  
DB 127 CYEDQGISYRGTWTSTASSGAECTNW--NSSALAKPYSGRRPDALRLGLGNHYCRNPDR 184  
QY 52 DPRGPWCYV 60  
DB 185 DSK-PWCYV 192  
RESULT 3

HGFA\_MOUSE  
ID AC HGFA\_MOUSE STANDARD; PRT; 653 AA.  
AD Q9R058; Q9JKV4;  
DT 16-OCT-2001 (Rel. 40, Created)  
DT 16-OCT-2001 (Rel. 40, Last sequence update)  
DT 15-MAR-2004 (Rel. 43, Last annotation update)  
DE Hepatocyte growth factor activator precursor (EC 3.4.21.-) (HGF  
activator) (HGFA).  
GN HGFA.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
PP SEQUENCE FROM N.A.  
RC STRAIN=BALE/C;  
RA Itoh H., Kataoka H., Koono H.;  
RL "Mouse hepatocyte growth factor activator";  
RL Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.  
[2]  
RN SEQUENCE FROM N.A.  
RX MEDLINE=21226753; PubMed=11032833;  
RA van Adelsberg J.S., Sehgal S., Kukes A., Brady C., Barasch J.,  
RA Yang J., Huang Y.;  
RT "Activation of hepatocyte growth factor (HGF) by endogenous HGF  
activator is required for mesangial kidney morphogenesis in  
vitro";  
RL J. Biol. Chem. 276:15099-15106 (2001).  
CC -!- FUNCTION: Activates hepatocyte growth factor (HGF) by converting  
it from a single chain to a heterodimeric form (By similarity).  
CC -!- SUBUNIT: Dimer of a short chain and a long chain linked by a  
disulfide bond (By similarity).  
CC -!- SUBCELLULAR LOCATION: Secreted as an inactive single-chain  
precursor and is then activated to a heterodimeric form (By  
similarity).  
CC -!- SIMILARITY: Belongs to peptidase family S1.  
CC -!- SIMILARITY: Contains 2 EGF-like domains.  
CC -!- SIMILARITY: Contains 1 fibronectin type I domain.  
CC -!- SIMILARITY: Contains 1 fibronectin type II domain.  
CC -!- SIMILARITY: Contains 1 kringle domain.  
-----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----  
DR EMBL; AF029017; AAF02489.1; -;  
DR EMBL; AF224724; AAF34712.1; -;  
DR HSSP; P00763; 1DPO.  
DR MEROPS; S01.228; -;  
DR MGD; MGI:1859281; Hgfac.  
DR InterPro; IPR009003; Cys\_ser\_trypsin.  
DR InterPro; IPR000742; EGF\_2.  
DR InterPro; IPR006209; EGF-like.  
DR InterPro; IPR000083; Fibrinctn1.  
DR InterPro; IPR000562; FN\_Type\_II.  
DR InterPro; IPR006210; IEGF.  
DR InterPro; IPR000001; Kringle.  
DR InterPro; IPR001254; Peptidase\_S1.  
DR InterPro; IPR001314; Peptidase\_S1A.  
DR Pfam; PF00008; EGF; 2.  
DR Pfam; PF00039; fn1; 1.  
DR Pfam; PF00040; fn2; 1.  
DR Pfam; PF00051; kringle; 1.  
DR Pfam; PF00089; trypsin; 1.  
DR PRINTS; PR00722; CHYMOTRYPSIN.  
DR PRINTS; PR00013; FNTYPEII.  
DR PRINTS; PR00018; KRINGLE.  
DR ProDom; PD000995; FN\_Type\_II; 1.  
DR ProDom; PD000395; Kringle; 1.

DR SMART; SM00181; EGF; 2.  
DR SMART; SM00059; FN2; 1.  
DR SMART; SM00130; KR; 1.  
DR SMART; SM00020; Tryp\_Spc; 1.  
DR PROSITE; PS00022; EGF\_1; 2.  
DR PROSITE; PS01186; EGF\_2; 1.  
DR PROSITE; PS00026; EGF\_3; 2.  
DR PROSITE; PS01253; FIBRONECTIN\_1; 1.  
DR PROSITE; PS00021; KRINGLE\_1; 1.  
DR PROSITE; PS00070; TRYPSIN\_DOM; 1.  
DR PROSITE; PS00240; TRYPSIN\_2; 1.  
DR PROSITE; PS00134; TRYPSIN\_HIS; 1.  
DR PROSITE; PS00135; TRYPSIN\_SER; 1.  
KW Hydrolase; Glycoprotein; Plasma; Serine protease; Kringle; Signal;  
KW EGF-like domain; Repeat; Zymogen.  
FT SIGNAL 1 29 BY SIMILARITY.  
FT PROPEP 30 369 CLEAVED IN ACTIVE FORM (BY SIMILARITY).  
FT CHAIN 370 405 HEPATOCYTE GROWTH FACTOR ACTIVATOR SHORT  
CHAIN.  
FT CHAIN 406 653 HEPATOCYTE GROWTH FACTOR ACTIVATOR LONG  
CHAIN.  
FT DOMAIN 105 145 FIBRONECTIN TYPE-II.  
FT DOMAIN 157 195 EGF-LIKE 1.  
FT DOMAIN 197 237 FIBRONECTIN TYPE-I.  
FT DOMAIN 238 276 EGF-LIKE 2.  
FT DOMAIN 283 364 KRINGLE.  
FT DOMAIN 406 653 SERINE PROTEASE.  
FT ACT\_SITE 445 445 CHARGE RELAY SYSTEM (BY SIMILARITY).  
FT ACT\_SITE 495 495 CHARGE RELAY SYSTEM (BY SIMILARITY).  
FT ACT\_SITE 596 596 CHARGE RELAY SYSTEM (BY SIMILARITY).  
FT DISULFID 105 130 BY SIMILARITY.  
FT DISULFID 119 145 BY SIMILARITY.  
FT DISULFID 161 172 BY SIMILARITY.  
FT DISULFID 166 183 BY SIMILARITY.  
FT DISULFID 185 194 BY SIMILARITY.  
FT DISULFID 199 227 BY SIMILARITY.  
FT DISULFID 225 234 BY SIMILARITY.  
FT DISULFID 242 253 BY SIMILARITY.  
FT DISULFID 247 264 BY SIMILARITY.  
FT DISULFID 266 275 BY SIMILARITY.  
FT DISULFID 283 364 BY SIMILARITY.  
FT DISULFID 304 346 BY SIMILARITY.  
FT DISULFID 335 359 BY SIMILARITY.  
FT DISULFID 392 519 INTERCHAIN (BY SIMILARITY).  
FT DISULFID 430 446 BY SIMILARITY.  
FT DISULFID 438 508 BY SIMILARITY.  
FT DISULFID 533 602 BY SIMILARITY.  
FT DISULFID 565 581 BY SIMILARITY.  
FT DISULFID 592 620 BY SIMILARITY.  
FT CARBOHYD 39 39 N-LINKED (GLCNAC... ) (POTENTIAL).  
FT CARBOHYD 47 47 N-LINKED (GLCNAC... ) (POTENTIAL).  
FT CARBOHYD 63 63 N-LINKED (GLCNAC... ) (POTENTIAL).  
FT CARBOHYD 287 287 N-LINKED (GLCNAC... ) (POTENTIAL).  
FT CARBOHYD 466 466 N-LINKED (GLCNAC... ) (POTENTIAL).  
FT CARBOHYD 544 544 N-LINKED (GLCNAC... ) (POTENTIAL).  
FT CONFLICT 164 164 G -> W (IN REF. 2).  
SQ SEQUENCE 653 AA; 70567 MW; 89B4B20255DF7FDC CRC64;  
Query Match 40.9%; Score 146; DB 1; Length 653;  
Best Local Similarity 47.1%; Pred. No. 3.2e-09;  
Matches 32; Conservative 19; Mismatches 10; Gaps 3;  
QY 1 CFWDNGHLYREDQTSAPGLRCINW-----LDAQSGLASAPVSGAGNSYCNPD 52  
DB 283 CFLNGTETRGVASTAAGSLCLAWNSDLLYOLHVDG-VAAAVLGLGPHYCRNPD 341  
QY 53 PRGFWCYV 60  
DB 342 ER-PWCYV 348

RESULT 4

TPA\_RAT ID TPA\_RAT STANDARD; PRT; 559 AA.

AC P1937;

DT 01-FEB-1991 (Rel. 17, Created)

DT 01-NOV-1991 (Rel. 20, Last sequence update)

DT 10-OCT-2003 (Rel. 42, Last annotation update)

DE Tissue-type plasminogen activator precursor (EC 3.4.21.68) (tPA)

DE (t-PA) (t-plasminogen activator).

GN PLAT.

OS Rattus norvegicus (Rat).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.

NCBI\_TaxID=10116;

RN [1]

RX MEDLINE=89170114; PubMed=3148445;

RA NY T., Leonardson G., Haueh A.J.W.;

RT "Cloning and characterization of a cDNA for rat tissue-type plasminogen activator.";

RL DNA 7:671-677(1988).

RN [2]

RX MEDLINE=90130448; PubMed=2105315;

RA Feng P., Ohlsson M., Ny T.;

RT "The structure of the TATA-less rat tissue-type plasminogen activator gene. Species-specific sequence divergences in the promoter predict differences in regulation of gene expression.";

RL J. Biol. Chem. 265:2022-2027(1990).

CC -!- FUNCTION: Converts the abundant, but inactive, zymogen plasminogen to plasmin by hydrolyzing a single Arg-Val bond in plasminogen. By controlling plasmin-mediated proteolysis, it plays an important role in tissue remodeling and degradation, in cell migration and many other physiological events.

CC -!- CATALYTIC ACTIVITY: Specific cleavage of Arg-|-Val bond in plasminogen to form plasmin.

CC -!- SUBUNIT: Heterodimer of chain A and chain B held by a disulfide bond.

CC -!- SUBCELLULAR LOCATION: Secreted; extracellular.

CC -!- PM: THE SINGLE CHAIN, ALMOST FULLY ACTIVE ENZYME, CAN BE FURTHER PROCESSED INTO A TWO-CHAIN FULLY ACTIVE FORM BY A CLEAVAGE AFTER ARG-308 CATALYZED BY PLASMIN, TISSUE KALLIKREIN OR FACTOR XA.

CC -!- MISCELLANEOUS: Binds to the kringle structure of the fibrin A chain. Binding to fibrin enhances its catalytic activity.

CC -!- SIMILARITY: Belongs to peptidase family S1.

CC -!- SIMILARITY: Contains 1 EGF-like domain.

CC -!- SIMILARITY: Contains 1 fibronectin type I domain.

CC -!- SIMILARITY: Contains 2 kringle domains.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR MEROPS; S01.232; -

DR InterPro; IPR009003; Cys\_Ser\_trypsin.

DR InterPro; IPR006209; EGF-like.

DR InterPro; IPR000083; Fibronctn1.

DR InterPro; IPR006210; IEGF.

DR InterPro; IPR000001; Kringle.

DR InterPro; IPR001254; Peptidase S1.

DR InterPro; IPR001314; Peptidase\_S1A.

DR Pfam; PF00008; EGF; 1.

DR Pfam; PF00039; fnl; 1.

DR Pfam; PF00051; kringle; 2.

DR Pfam; PF00089; trypsin; 1.

DR PRINTS; PR00722; CHYMOTRYPSIN.

DR PRINTS; PR00018; KRINGLE.

DR PRODOM; PD000395; Kringle; 2.

DR SMART; SM00381; EGF; 1.

DR SMART; SM00058; FNI; 1.

DR SMART; SM00130; KR; 2.

DR SMART; SM00020; Tryp\_Spc; 1.

DR PROSITE; PS00022; EGF\_1; 1.

DR PROSITE; PS01186; EGF\_2; 1.

DR PROSITE; PS00025; EGF\_3; 1.

DR PROSITE; PS01253; FIBRONECTIN\_1; 1.

DR PROSITE; PS00021; KRINGLE\_1; 2.

DR PROSITE; PS00070; KRINGLE\_2; 2.

DR PROSITE; PS0240; TRYPSIN\_DOM; 1.

DR PROSITE; PS00134; TRYPSIN\_HIS; 1.

DR PROSITE; PS00135; TRYPSIN\_SER; 1.

KW Plasminogen activation; Hydrolase; Serine protease; Glycoprotein; Plasma; Kringle; EGF-like domain; Repeat; Signal.

KW SIGNAL

FT PROPEP 18 29

FT CHAIN 30 559

FT CHAIN 30 308

FT CHAIN 309 559

FT CHAIN 36 78

FT DOMAIN 79 117

FT DOMAIN 124 205

FT DOMAIN 213 294

FT DOMAIN 309 559

FT ACT\_SITE 355 355

FT ACT\_SITE 404 404

FT ACT\_SITE 510 510

FT DISULFID 38 68

FT DISULFID 66 75

FT DISULFID 83 94

FT DISULFID 88 105

FT DISULFID 107 116

FT DISULFID 124 205

FT DISULFID 145 187

FT DISULFID 176 200

FT DISULFID 213 294

FT DISULFID 234 276

FT DISULFID 265 289

FT DISULFID 297 428

FT DISULFID 340 356

FT DISULFID 348 417

FT DISULFID 442 516

FT DISULFID 474 490

FT DISULFID 506 534

FT CARBOHYD 149 149

FT CARBOHYD 481 481

FT CONFLICT 380 380

SQ SEQUENCE 559 AA; 62903 MW; 7DBD3809C1D1C921 CRC64;

Query Match 39.5%; Score 141; DB 1; Length 559;

Best Local Similarity 42.0%; Pred. No. 1e-08; 22; Indels 12; Gaps 3;

Matches 29; Conservative 6; Mismatches

QY 1 CFWDNGHLYREDQTSAPGLRCINLWLDQSGLASAPVS-----GAGNHSYCRNPDE 51

Db 124 CPEGQITVGTWSTAENGACINW--NSSALSQKPYSGARRPNAIKLGLGNHNYCRNPD 181  
QY 52 DPGPGWCYV 60  
Db 182 DVK-PWCYV 189

RESULT 5  
HGFA HUMAN STANDARD; PRT; 655 AA.  
AC Q04756; Q14726;  
DT 01-JUN-1994 (Rel. 29, Created)  
DT 01-JUN-1994 (Rel. 29, Last sequence update)  
DT 10-OCT-2003 (Rel. 42, Last annotation update)  
DE Hepatocyte growth factor activator precursor (EC 3.4.21.-) (HGF  
DE activator) (HGFA).  
GN HGFA.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.  
RC TISSUE=Liver, and Serum;  
RX MEDLINE=93252878; PubMed=7683665;  
RA Miyazawa K., Shimomura T., Kitamura A., Kondo J., Morimoto Y.,  
RA Kitamura N.;  
RT "Molecular cloning and sequence analysis of the cDNA for a human  
RT serine protease responsible for activation of hepatocyte growth  
RT factor. Structural similarity of the protease precursor to blood  
RT coagulation factor XII.";  
RL J. Biol. Chem. 268:10024-10028(1993).  
RN [2]  
RP SEQUENCE OF 40-555 FROM N.A.  
RA Zhao S., Odell C.;  
RL Submitted (FEB-1996) to the EMBL/GenBank/DBJ databases.  
CC -!- FUNCTION: Activates hepatocyte growth factor (HGF) by  
CC converting it from a single chain to a heterodimeric form.  
CC -!- SUBUNIT: Dimer of a short chain and a long chain linked by a  
CC disulfide bond.  
CC -!- SUBCELLULAR LOCATION: Secreted as an inactive single-chain  
CC precursor and is then activated to a heterodimeric form.  
CC -!- TISSUE SPECIFICITY: Liver.  
CC -!- SIMILARITY: Belongs to peptidase family S1.  
CC -!- SIMILARITY: Contains 2 EGF-like domains.  
CC -!- SIMILARITY: Contains 1 fibronectin type I domain.  
CC -!- SIMILARITY: Contains 1 fibronectin type II domain.  
CC -!- SIMILARITY: Contains 1 kringle domain.  
CC -!- SIMILARITY: Contains 1 kringle domain.  
CC -!- CAUTION: It is uncertain whether Met-1 is the initiator.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
CC  
CC EMBL; D14012; BAA03113.1; -;  
CC EMBL; Z69923; -; NOT\_ANNOTATED\_CDS.  
CC PIR; A46688; A46688.  
CC HSRP; P00763; LDPO.  
CC MEROPS; S01.228; -;  
CC Genew; HGNC:4894; HGFA.  
CC MIM; 604552; -;  
CC GO; GO:0005576; C:extracellular; TAS.  
CC GO; GO:004252; P:serine-type endopeptidase activity; TAS.  
CC GO; GO:006508; P:proteolysis and peptidolysis; TAS.  
CC InterPro; IPR009003; Cys\_Ser\_trypsin.  
CC InterPro; IPR000742; EGF\_2.  
CC InterPro; IPR006209; EGF-like.  
CC InterPro; IPR000083; Fibrinctnl.  
CC InterPro; IPR000562; FN\_Type\_II.

DR InterPro; IPR006210; IEGF.  
DR InterPro; IPR000001; Kringle.  
DR InterPro; IPR001254; Peptidase\_S1.  
DR InterPro; IPR001314; Peptidase\_S1A.  
DR Pfam; PF00038; EGF\_2.  
DR Pfam; PF00039; fn1; 1.  
DR Pfam; PF00040; fn2; 1.  
DR Pfam; PF00051; kringle; 1.  
DR Pfam; PF00089; trypsin; 1.  
DR PRINTS; PR00722; CHYMOTRYPSIN.  
DR PRINTS; PR00013; FNTYPEII.  
DR PRINTS; PR00018; KRINGLE.  
DR ProDom; PD000995; FN\_Type\_II; 1.  
DR ProDom; PD000395; Kringle; 1.  
DR SMART; SM00181; EGF\_2.  
DR SMART; SM00058; FN1; 1.  
DR SMART; SM00059; FN2; 1.  
DR SMART; SM00130; KR; 1.  
DR SMART; SM00200; Tryp\_SPC; 1.  
DR PROSITE; PS00022; EGF\_1; 2.  
DR PROSITE; PS01186; EGF\_2; 1.  
DR PROSITE; PS00026; EGF\_3; 2.  
DR PROSITE; PS01253; FIBRONECTIN\_1; 1.  
DR PROSITE; PS00021; FIBRONECTIN\_2; 1.  
DR PROSITE; PS00021; KRINGLE\_1; 1.  
DR PROSITE; PS00070; KRINGLE\_2; 1.  
DR PROSITE; PS50024; TRYPsin\_DOM; 1.  
DR PROSITE; PS00134; TRYPsin\_HIS; 1.  
DR PROSITE; PS00135; TRYPsin\_SER; 1.  
KW Hydrolase; Glycoprotein; Plasma; Serine protease; Kringle; Signal;  
KW EGF-like domain; Repeat; Zymogen.  
FT SIGNAL 1 30  
FT PROPEP 31 372  
FT CHAIN 373 407  
FT CHAIN 408 655  
FT CHAIN 655 655  
FT DOMAIN 108 148  
FT DOMAIN 160 198  
FT DOMAIN 200 240  
FT DOMAIN 241 279  
FT DOMAIN 286 367  
FT DOMAIN 408 655  
FT ACT\_SITE 447 447  
FT ACT\_SITE 497 497  
FT ACT\_SITE 598 598  
FT DISULFID 108 133  
FT DISULFID 122 148  
FT DISULFID 164 175  
FT DISULFID 169 186  
FT DISULFID 188 197  
FT DISULFID 202 230  
FT DISULFID 228 237  
FT DISULFID 245 256  
FT DISULFID 250 267  
FT DISULFID 269 278  
FT DISULFID 286 367  
FT DISULFID 307 349  
FT DISULFID 338 362  
FT DISULFID 394 521  
FT DISULFID 432 448  
FT DISULFID 440 510  
FT DISULFID 535 604  
FT DISULFID 567 583  
FT DISULFID 594 622  
FT CARBOHYD 48 48  
FT CARBOHYD 290 290  
FT CARBOHYD 468 468  
FT CARBOHYD 492 492  
FT CARBOHYD 546 546  
FT CARBOHYD 644 644  
FT CONFLICT 655 AA; 70681 MW; 2CF72F1E1B862ED7 CRC64;  
SQ SEQUENCE



FT DISULFID 234 276 BY SIMILARITY.  
 FT DISULFID 265 289 BY SIMILARITY.  
 FT DISULFID 297 428 INTERCHAIN (BY SIMILARITY).  
 FT DISULFID 340 356 BY SIMILARITY.  
 FT DISULFID 348 417 BY SIMILARITY.  
 FT DISULFID 442 516 BY SIMILARITY.  
 FT DISULFID 474 490 BY SIMILARITY.  
 FT DISULFID 506 534 BY SIMILARITY.  
 FT CARBOHYD 149 139 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 481 481 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CONFLICT 260 260 G -> A (IN REF. 1).  
 FT CONFLICT 325 325 P -> A (IN REF. 1).  
 SQ SEQUENCE 559 AA; 63122 MW; 8CCE2BDB94514D9 CRC64;

Query Match 38.1%; Score 136; DB 1; Length 559;  
 Best Local Similarity 40.6%; Pred. No. 3.8e-08;  
 Matches 28; Conservative 8; Mismatches 21; Indels 12; Gaps 3;

QY 1 CFWDNHLYREDQTSAPGLRLNWLDAQSLGASAPVS-----GAGNHSYCRNPD 51  
 124 CFEQGITVGTWTSAESGAECINW--NSVLSLKPYNARRRPAKLGNGHNYCRNPD 181

QY 52 DPRGPWCYV 60  
 182 DLK-PWCYV 189

RESULT 7  
 ID URT2\_DESRO STANDARD; PRT; 431 AA.  
 AC P98121; 1996 (Rel. 33, Created)  
 DT 01-FEB-1996 (Rel. 33, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Salivary plasminogen activator beta precursor (EC 3.4.21.68) (DSPA beta)  
 OS Desmodus rotundus (Vampire bat).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Euthera; Chiroptera; Microchiroptera; Phyllostomidae;  
 OC Desmodontinae; Desmodus.  
 OX NCBI\_TaxID=9430;  
 RN [1]  
 RN SEQUENCE FROM N.A.  
 RP TISSUE=Salivary gland;  
 RX MEDLINE=92039036; PubMed=1937019;  
 RA Kraetzschmar J., Haendler B., Langer G., Boidol W., Bringmann P., Alagon A., Donner P., Schleuning W.D.;  
 RT "The plasminogen activator family from the salivary gland of the vampire bat Desmodus rotundus: cloning and expression.";  
 RL Gene 105:229-237(1991).  
 RN [2]  
 RP CHARACTERIZATION.  
 RX MEDLINE=9393059; PubMed=1309059;  
 RA Schleuning W.-D., Alagon A., Boidol W., Bringmann P., Petri T., Kraetzschmar J., Haendler B., Langer G., Baldus B., Witt W., Donner P.;  
 RT "Plasminogen activators from the saliva of Desmodus rotundus (common vampire bat): unique fibrin specificity.";  
 RL Ann. N.Y. Acad. Sci. 667:395-403(1992)  
 CC -!- FUNCTION. Probably essential to support the feeding habits of this exclusively haematophagous animal. Probable potent thrombolytic agent.  
 CC -!- CATALYTIC ACTIVITY. Specific cleavage of Arg-|-Val bond in plasminogen to form plasmin.  
 CC -!- SUBUNIT. Monomer.  
 CC -!- SIMILARITY. Belongs to peptidase family S1.  
 CC -!- SIMILARITY. Contains 1 EGF-like domain.  
 CC -!- SIMILARITY. Contains 1 kringle domain.

-----  
 This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).  
 CC -----  
 CC EMBL; M63989; AAA31594.1; -  
 CC PIR; J50599; J50599.  
 CC HSP; P98119; IASI.  
 CC MEROPS; S01.239; -  
 CC InterPro; IPR009003; Cys ser trypsin.  
 CC InterPro; IPR006209; EGF-like.  
 CC InterPro; IPR006210; IEGF.  
 CC InterPro; IPR000001; Kringle.  
 CC InterPro; IPR001254; Peptidase S1.  
 CC InterPro; IPR001314; Peptidase\_S1A.  
 CC Pfam; PF00008; EGF\_1.  
 CC Pfam; PF00051; kringle; 1.  
 CC Pfam; PF00089; trypsin; 1.  
 CC PRINTS; PR00722; CHYMOTRYPSIN.  
 CC PRINTS; PR00018; KRINGLE.  
 CC PRODOM; PD000395; Kringle; 1.  
 CC SMART; SMO0181; EGF\_1.  
 CC SMART; SMO0130; KR; 1.  
 CC SMART; SMO0020; TYD\_SPC; 1.  
 CC PROSITE; PS00022; EGF\_1; 1.  
 CC PROSITE; PS01186; EGF\_2; 1.  
 CC PROSITE; PS00028; EGF\_3; 1.  
 CC PROSITE; PS00021; KRINGLE\_1; 1.  
 CC PROSITE; PS00070; KRINGLE\_2; 1.  
 CC PROSITE; PS02040; TRYPSIN\_DOM; 1.  
 CC PROSITE; PS00134; TRYPSIN\_HIS; 1.  
 CC PROSITE; PS00135; TRYPSIN\_SER; 1.  
 CC Plasminogen activation; Hydrolase; Serine protease; Glycoprotein;  
 KW Kringle; EGF-like domain; Signal; Multigene family.  
 KW SIGNAL 36  
 FT CHAIN 37 431 SALIVARY PLASMINOGEN ACTIVATOR BETA.  
 FT DOMAIN 37 75 EGF-LIKE.  
 FT DOMAIN 82 163 KRINGLE.  
 FT DOMAIN 179 431 SERINE PROTEASE.  
 FT ACT\_SITE 226 226 CHARGE RELAY SYSTEM (BY SIMILARITY).  
 FT ACT\_SITE 275 275 CHARGE RELAY SYSTEM (BY SIMILARITY).  
 FT ACT\_SITE 382 382 CHARGE RELAY SYSTEM (BY SIMILARITY).  
 FT DISULFID 41 52 BY SIMILARITY.  
 FT DISULFID 46 63 BY SIMILARITY.  
 FT DISULFID 65 74 BY SIMILARITY.  
 FT DISULFID 82 163 BY SIMILARITY.  
 FT DISULFID 103 145 BY SIMILARITY.  
 FT DISULFID 134 158 BY SIMILARITY.  
 FT DISULFID 168 299 BY SIMILARITY.  
 FT DISULFID 211 227 BY SIMILARITY.  
 FT DISULFID 219 288 BY SIMILARITY.  
 FT DISULFID 313 388 BY SIMILARITY.  
 FT DISULFID 345 361 BY SIMILARITY.  
 FT DISULFID 378 406 BY SIMILARITY.  
 FT CARBOHYD 139 139 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 352 352 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 SQ SEQUENCE 431 AA; 48221 MW; 699B5E675B162CBF CRC64;

Query Match 37.5%; Score 134; DB 1; Length 431;  
 Best Local Similarity 38.6%; Pred. No. 5e-08;  
 Matches 27; Conservative 9; Mismatches 20; Indels 14; Gaps 3;

QY 1 CFWDNHLYREDQTSAPGLRLNWLDAQSLGASAPVS-----GAGNHSYCRNPD 50  
 82 CYKQGVYGTWTSTSEGAOCINW--NSNLLRTYNGRRSDAILGLGNHNYCRNPD 138

QY 51 EDPRGPWCYV 60  
 139 NNSK-PWCYV 147

RESULT 8  
 URT2\_DESRO STANDARD; PRT; 477 AA.  
 ID URT2\_DESRO STANDARD; PRT; 477 AA.

AC PI5638;  
 DT 01-APR-1990 (Rel. 14, Created)  
 DT 01-FEB-1996 (Rel. 33, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Salivary plasminogen activator alpha 2 precursor (EC 3.4.21.68) (DSPA  
 DE alpha-2) (BAT-2A) (T-plasminogen activator).  
 OS Desmodus rotundus (Vampire bat).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Chiroptera; Microchiroptera; Phyllostomidae;  
 OC Desmodontinae; Desmodus.  
 OX NCBI\_TaxID=9430;  
 RN [1]  
 RN SEQUENCE FROM N.A.  
 RC TISSUE=Salivary Gland;  
 RX MEDLINE=92039036; PubMed=1937019;  
 RA Kraetzschmar J., Haendler B., Langer G., Boidol W., Bringmann P.,  
 RA Alagon A., Donner P., Schleuning W.D.;  
 RT "The plasminogen activator family from the salivary gland of the  
 RT vampire bat Desmodus rotundus: cloning and expression.";  
 RL Gene 105:229-237(1991).  
 RN [2]  
 RN SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.  
 RC TISSUE=Salivary gland;  
 RX MEDLINE=9003687; PubMed=2509450;  
 RA Gardell S.J., Duong L.T., Diehl R.E., York J.D., Hare T.R.,  
 RA Register R.B., Jacobs J.W., Dixon R.A.F., Friedman P.A.;  
 RT "Isolation, characterization, and cDNA cloning of a vampire bat  
 RT salivary plasminogen activator.";  
 RL J. Biol. Chem. 264:17947-17952(1989).  
 RN [3]  
 RN CHARACTERIZATION.  
 RX MEDLINE=93393059; PubMed=1309059;  
 RA Schleuning W.-D., Alagon A., Boidol W., Bringmann P., Petri T.,  
 RA Kraetzschmar J., Haendler B., Langer G., Baldus B., Witt W.,  
 RA Donner P.;  
 RT "Plasminogen activators from the saliva of Desmodus rotundus (common  
 RT vampire bat): unique fibrin specificity.";  
 RL Ann. N.Y. Acad. Sci. 667:395-403(1992).  
 CC -/- FUNCTION: Probably essential to support the feeding habits of this  
 CC exclusively haematophagous animal. Probable potent thrombolytic  
 CC agent.  
 CC -/- CATALYTIC ACTIVITY: Specific cleavage of Arg-|-val bond in  
 CC plasminogen to form plasmin.  
 CC -/- ENZYME REGULATION: Activity toward plasminogen is stimulated in  
 CC the presence of fibrin I.  
 CC -/- SUBUNIT: Monomer.  
 CC -/- DOMAIN: The fibronectin type-I domain mediates binding to fibrin,  
 CC and the kringle domain apparently mediates fibrin-induced  
 CC stimulation of activity.  
 CC -/- SIMILARITY: Belongs to peptidase family S1.  
 CC -/- SIMILARITY: Contains 1 EGF-like domain.  
 CC -/- SIMILARITY: Contains 1 fibronectin type I domain.  
 CC -/- SIMILARITY: Contains 1 kringle domain.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; M63988; AAA31593.1; --  
 DR EMBL; J05082; AAA31596.1; --  
 DR PIR; A34369; A34369.  
 DR PIR; J05098; J05098.  
 DR HSP; P98119; IAS1.  
 DR MEROPS; S01.232; --  
 DR InterPro; IPR009003; Cys Ser trypsin.  
 DR InterPro; IPR006209; EGF-like.  
 DR InterPro; IPR000083; Fibinctnl.  
 DR InterPro; IPR006210; IEGF.  
 DR InterPro; IPR000001; Kringle.

DR InterPro; IPR001254; Peptidase S1.  
 DR InterPro; IPR001314; Peptidase\_S1A.  
 DR Pfam; PF00008; EGF\_1.  
 DR Pfam; PF00039; fn1\_1.  
 DR Pfam; PF00051; Kringle\_1.  
 DR Pfam; PF00085; trypsin\_1.  
 DR PRINTS; PR00722; CHIMOTRYPSIN.  
 DR PRINTS; PR00018; KRINGLE.  
 DR ProDom; PD000395; Kringle\_1.  
 DR SMART; SM00181; EGF\_1.  
 DR SMART; SM00058; FN1\_1.  
 DR SMART; SM00130; KR\_1.  
 DR SMART; SM00020; Tryp\_SPC\_1.  
 DR PROSITE; PS00022; EGF\_1; 1.  
 DR PROSITE; PS01186; EGF\_2; 1.  
 DR PROSITE; PS00026; EGF\_3; 1.  
 DR PROSITE; PS01253; FIBRONECTIN\_1; 1.  
 DR PROSITE; PS00021; KRINGLE\_1; 1.  
 DR PROSITE; PS00070; KRINGLE\_2; 1.  
 DR PROSITE; PS02040; TRYPSIN\_DOM; 1.  
 DR PROSITE; PS00134; TRYPSIN\_HIS; 1.  
 DR PROSITE; PS00135; TRYPSIN\_SER; 1.  
 DR Plasminogen activation; Hydrolase; Serine protease; Glycoprotein;  
 KW Kringle; EGF-like domain; Signal; Multigene family.  
 FT SIGNAL 1 36  
 FT CHAIN 37 477  
 FT DOMAIN 40 82  
 FT DOMAIN 83 121  
 FT DOMAIN 128 209  
 FT DOMAIN 225 477  
 FT ACT SITE 272 272  
 FT ACT SITE 321 321  
 FT ACT SITE 428 428  
 FT DISULFID 42 72  
 FT DISULFID 70 79  
 FT DISULFID 87 98  
 FT DISULFID 92 109  
 FT DISULFID 111 120  
 FT DISULFID 128 209  
 FT DISULFID 149 191  
 FT DISULFID 180 204  
 FT DISULFID 214 345  
 FT DISULFID 257 273  
 FT DISULFID 265 334  
 FT DISULFID 359 434  
 FT DISULFID 391 407  
 FT DISULFID 424 452  
 FT CARBOHYD 185 185  
 FT CARBOHYD 398 398  
 FT CONFLICT 403 403  
 FT CONFLICT 417 417  
 FT CONFLICT 435 435  
 SQ SEQUENCE 477 AA; 53719 MW; 17486555C0E5077C CRC64;  
 Query Match 37.5%; Score 134; DB 1; Length 477;  
 Best Local Similarity 38.6%; Pred. No. 5 Se-09;  
 Matches 27; Conservative 9; Mismatches 20; Indels 14; Gaps 3;  
 QY 1 CFWNGHLYREDQTSPPAGRLCNLWDAQSL-----ASAPVSGAGNHSYCRNPD 50  
 128 CYKQGVYTRGTSTSSGAQCNW---NSNLLTRTYNGRRSDAITLGLGNHNYCRNPD 184  
 QY 51 EDPRGPWCYV 60  
 185 NNSK-PWCYV 193  
 Db  
 RESULT 9  
 ID UROK\_PAPCY STANDARD; PRT; 433 AA.  
 AC P16227;  
 DT 01-APR-1990 (Rel. 14, Created)  
 DT 01-APR-1990 (Rel. 14, Last sequence update)



DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE Urokinase-type plasminogen activator precursor (EC 3.4.21.73) (uPA)  
DE (U-plasminogen activator)  
GN PLAU.  
OS Papio cynocephalus (Yellow baboon).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea;  
OC Cercopithecoidea; Papio.  
OX NCBI\_TaxID=9556;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=thoracic aorta;  
RA MEDLINE=90287734; PubMed=2113276;  
RA Au Y.P.T., Wang T.W., Clowes A.W.;  
RT "Nucleotide and deduced amino acid sequences of baboon urokinase-type  
RL plasmminogen activator.";  
RL Nucleic Acids Res. 18:3411-3411(1990).  
CC -1- CATALYTIC ACTIVITY: Specific cleavage of Arg-|-Val bond in  
CC plasmminogen to form plasmin.  
CC -1- SUBUNIT: Found in high and low molecular mass forms. Each consists  
CC of two chains, A and B. The high molecular mass form contains a  
CC long chain A. Cleavage occurs after residue 155 in the low  
CC molecular mass form to yield a short A1 chain (By similarity).  
CC -1- SIMILARITY: Belongs to peptidase family S1.  
CC -1- SIMILARITY: Contains 1 EGF-like domain.  
CC -1- SIMILARITY: Contains 1 kringle domain.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
CC at the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC DR EMBL; X51935; CAA36200.1; --  
CC DR PIR; S14687; URBA.  
CC DR HSP; P00749; ILMW.  
CC DR MEROPS; S01.231; --  
CC DR InterPro; IPR009003; Cys Ser trypsin.  
CC DR InterPro; IPR006203; EGF-like.  
CC DR InterPro; IPR006210; IEGF.  
CC DR InterPro; IPR000001; Kringle.  
CC DR InterPro; IPR008293; Pept S1A uPA.  
CC DR InterPro; IPR001254; Peptidase S1.  
CC DR InterPro; IPR001314; Peptidase\_S1A.  
CC DR Pfam; PF00051; kringle; 1.  
CC DR Pfam; PF00039; trypsin; 1.  
CC DR PfamSF; PIRSF001144; Urokinase; 1.  
CC DR PRINTS; PR00722; CHYMOTRYPSIN.  
CC DR PRINTS; PR00018; KRINGLE.  
CC DR ProDom; PD000395; Kringle; 1.  
CC DR SMART; SM00181; EGF; 1.  
CC DR SMART; SM00130; KR; 1.  
CC DR SMART; SM00020; Tryp\_SPC; 1.  
CC DR PROSITE; PS00022; EGF\_1; 1.  
CC DR PROSITE; PS0186; EGF\_2; FALSE\_NEG.  
CC DR PROSITE; PS00026; EGF\_3; 1.  
CC DR PROSITE; PS00021; KRINGLE\_1; 1.  
CC DR PROSITE; PS00070; KRINGLE\_2; 1.  
CC DR PROSITE; PS00240; TRYPsin\_DOM; 1.  
CC DR PROSITE; PS00134; TRYPsin\_HIS; 1.  
CC DR PROSITE; PS00135; TRYPsin\_SER; 1.  
CC DR Plasmminogen activation; Hydrolase; Serine protease; Glycoprotein;  
CC Kringle; EGF-like domain; Zymogen; Signal.  
CC SIGNAL 1 20 POTENTIAL.  
CC CHAIN 21 433 UROKINASE-TYPE PLASMINOGEN ACTIVATOR.  
CC FT CHAIN 21 176 CHAIN A (BY SIMILARITY).  
CC FT CHAIN 155 176 SHORT A CHAIN (A1) (BY SIMILARITY).  
CC FT CHAIN 178 433 CHAIN B (BY SIMILARITY).  
CC FT DOMAIN 26 62 EGF-LIKE.  
CC FT DOMAIN 69 150 EGF-LIKE.  
CC FT DOMAIN 151 177 CONNECTING PEPTIDE.  
CC FT DOMAIN

FT DOMAIN 178 433 SERINE PROTEASE.  
FT DISULFID 30 38 BY SIMILARITY.  
FT DISULFID 32 50 BY SIMILARITY.  
FT DISULFID 52 61 BY SIMILARITY.  
FT DISULFID 167 298 INTERCHAIN (BY SIMILARITY).  
FT DISULFID 208 224 BY SIMILARITY.  
FT DISULFID 216 287 BY SIMILARITY.  
FT DISULFID 315 394 BY SIMILARITY.  
FT DISULFID 347 363 BY SIMILARITY.  
FT DISULFID 374 402 BY SIMILARITY.  
FT ACT\_SITE 223 223 CHARGE RELAY SYSTEM.  
FT ACT\_SITE 274 274 CHARGE RELAY SYSTEM.  
FT ACT\_SITE 378 378 CHARGE RELAY SYSTEM.  
FT CARBOHYD 324 324 N-LINKED (GLCNAC... (BY SIMILARITY).  
SQ SEQUENCE 433 AA; 48595 MW; 816D22DFEDC8792 CRC64;  
Query Match 37.0%; Score 132; DB 1; Length 433;  
Best Local Similarity 40.3%; Pred. No. 8.5e-08;  
Matches 27; Conservative 7; Mismatches 25; Indels 8; Gaps 2;  
Oy 1 CFWDNGHLIREDOTSPAPGLRCLNWLDA-----QSGLASAPVSGAGNHSYCNEDDP 53  
Db 69 CYEGNGHYFGKASTDTMGSRCLANSAATVLOQTYHAHRSALQGLGKHNYCRNPD-NR 127  
Oy 54 RGPWCYV 60  
Db 128 RRPWCYV 134  
RESULT 10  
UROK\_HUMAN STANDARD; PRT; 431 AA.  
AC P00749; Q15844; Q16618; Q969W6;  
DT 21-JUN-1986 (Rel. 01, Created)  
DT 20-MAR-1987 (Rel. 04, Last sequence update)  
DT 10-OCT-2003 (Rel. 42, Last annotation update)  
DE Urokinase-type plasminogen activator precursor (EC 3.4.21.73) (uPA)  
GN PLAU.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=85215647; PubMed=2987867;  
RX Riccio A., Grimaldi G., Verde P., Sebastio G., Boast S., Blasi F.;  
RT "The human urokinase-plasminogen activator gene and its promoter.";  
RL Nucleic Acids Res. 13:2759-2771(1985).  
RN [2]  
RP SEQUENCE FROM N.A.  
RA Holmes W.B., Pennica D., Blaber M., Rey M.W., Guenzler W.A.,  
RA Steffens G.J., Heyneker H.L.;  
RT "Cloning and expression of the gene for pro-urokinase in Escherichia  
RT coli.";  
RL Biotechnology 3:923-929(1985).  
RN [3]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=86058554; PubMed=2415429;  
RX Nagai M., Hiramatsu R., Kaneda T., Hayasuke N., Arimura H.,  
RA Nishida M., Suyama T.;  
RT "Molecular cloning of cDNA coding for human preprourokinase.";  
RN Gene 36:183-188(1985).  
RN [4]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=85203359; PubMed=3888571;  
RX Jacobs P., Cravador A., Loriau R., Brockly F., Colau B., Chuchana P.,  
RA van Elsen A., Herzog A., Bollen A.;  
RT "Molecular cloning, sequencing, and expression in Escherichia coli of  
RT human preprourokinase cDNA.";  
RL DNA 4:139-146(1985).  
RN [5]  
RP SEQUENCE FROM N.A.

RA Rieder M.J., Carrington D.P., Chung M.-W., Lee K.L., Poel C.L., Yi Q.,  
RA Nickerson D.A.;  
RA Submitted [JUN-2001] to the EMBL/GenBank/DDBJ databases.  
RN [6]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Lung;  
RX MEDLINE=22388257; PubMed=12477932;  
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,  
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,  
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,  
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,  
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,  
RA Brownstein M.J., Ustin T.B., Teshiyuki S., Carninci P., Prange C.,  
RA Rana S.A., Loquellano N.A., Peters G.J., Abramson R.D., Mullany S.J.,  
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,  
RA Villalon D.K., Murry D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
RA Fahy J., Helton E., Ketteman M., Madan A., Rodrigues S., Sanchez A.,  
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,  
RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smalilus D.E.,  
RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;  
RA "Generation and initial analysis of more than 15,000 full-length  
RT human and mouse cDNA sequences";  
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).  
RN [7]  
RP SEQUENCE OF 66-431 FROM N.A.  
RX MEDLINE=84272706; PubMed=6589620;  
RA Verde P., Stoppelli M.P., Galeffi P., di Nocera P., Blasi F.;  
RT "Identification and primary sequence of an unspliced human urokinase  
RT poly(A)+ RNA";  
RL Proc. Natl. Acad. Sci. U.S.A. 81:4727-4731(1984).  
RN [8]  
RP SEQUENCE OF 21-177  
RX MEDLINE=83055084; PubMed=6754569;  
RA Gunzler W.A., Steffens G.-J., Otting F., Kim S.-M.A., Frankus E.,  
RA Flohe L.;  
RT "The primary structure of high molecular mass urokinase from human  
RT urine. The complete amino acid sequence of the A chain";  
RL Hoppe-Seyler's Z. Physiol. Chem. 363:1155-1165(1982).  
RN [9]  
RP SEQUENCE OF 156-176 AND 179-224.  
RX MEDLINE=83003608; PubMed=6749491;  
RA Schaller J., Nick H., Rickli E.E., Gillesen D., Lergier W.,  
RA Studer R.O.;  
RT "Human low-molecular-weight urinary urokinase. Partial  
RT characterization and preliminary sequence data of the two polypeptide  
RT chains";  
RL Eur. J. Biochem. 125:251-257(1982).  
RN [10]  
RP SEQUENCE OF 158-410.  
RX MEDLINE=83055099; PubMed=6754572;  
RA Steffens G.J., Gunzler W.A., Otting F., Frankus E., Flohe L.;  
RT "The complete amino acid sequence of low molecular mass urokinase  
RT from human urine";  
RL Hoppe-Seyler's Z. Physiol. Chem. 363:1043-1058(1982).  
RN [11]  
RP X-RAY CRYSTALLOGRAPHY (2.5 ANGSTROMS).  
RX MEDLINE=96000858; PubMed=8591045;  
RA Spraggon G., Phillips C., Nowak U.K., Ponting C.P., Saunders D.,  
RA Dobson C.M., Stuart D.I., Jones E.Y.;  
RT "The crystal structure of the catalytic domain of human  
RT urokinase-type plasminogen activator";  
RL Structure 3:681-691(1995).  
RN [12]  
RP X-RAY CRYSTALLOGRAPHY (1.8 ANGSTROMS) OF 159-411.  
RX MEDLINE=20266327; PubMed=10805774;  
RA Sperl S., Jacob U., Arroyo de Prada N., Sturzebecher J., Wilhelm O.G.,  
RA Bode W., Magdolen V., Huber R., Moroder L.;  
RT "(4-aminomethyl)phenylguanidine derivatives as nonpeptidic highly  
RT selective inhibitors of human urokinase";

Proc. Natl. Acad. Sci. U.S.A. 97:5113-5118(2000).  
RN [13]  
RP STRUCTURE BY NMR.  
RX MEDLINE=89127526; PubMed=2536903;  
RA Oswald R.E., Bogusky M.J., Bamberger M., Smith R.A.G., Dobson C.M.;  
RT "Dynamics of the multidomain fibrinolytic protein urokinase from two-  
RT dimensional NMR";  
RL Nature 337:579-582(1989).  
RN [14]  
RP STRUCTURE BY NMR OF 67-155.  
RX MEDLINE=93003110; PubMed=1327118;  
RA Li X., Smith R.A.G., Dobson C.M.;  
RT "Sequential 1H NMR assignments and secondary structure of the kringle  
RT domain from urokinase";  
RL Biochemistry 31:9562-9571(1992).  
RN [15]  
RP STRUCTURE BY NMR OF 67-155.  
RX MEDLINE=94149701; PubMed=8107091;  
RA Li X., Bokman A.M., Llinas M., Smith R.A.G., Dobson C.M.;  
RT "Solution structure of the kringle domain from urokinase-type  
RT plasminogen activator";  
RL J. Mol. Biol. 235:1548-1559(1994).  
RN [16]  
RP VARIANT LEU-141.  
RX MEDLINE=96186279; PubMed=8652631;  
RA Yoshimoto M., Ushiyama Y., Sakai M., Tamaki S., Hara H., Takahashi K.,  
RA Sawasaki Y., Hanada K.;  
RT "Characterization of single chain urokinase-type plasminogen  
RT activator with a novel amino-acid substitution in the kringle  
RT structure";  
RL Biochim. Biophys. Acta 1293:83-89(1996).  
RN [17]  
RP VARIANT LEU-141.  
RX MEDLINE=97218551; PubMed=9065988;  
RA Conne B., Berczy M., Belin D.;  
RT "Detection of polymorphisms in the human urokinase-type plasminogen  
RT activator gene";  
RL Thromb. Haemost. 77:434-435(1997).  
RN [18]  
RP ERRATUM.  
RA Conne B., Berczy M., Belin D.;  
RL Thromb. Haemost. 78:973-973(1997).  
RN [19]  
RP VARIANT LEU-141.  
RX MEDLINE=97337920; PubMed=9194591;  
RA Turkmen B., Schmitt M., Schmalfeldt B., Trommler P., Hell W.,  
RA Creutzburg S., Graeff H., Magdolen V.;  
RT "Mutational analysis of the genes encoding urokinase-type plasminogen  
RT activator (uPA) and its inhibitor PAI-1 in advanced ovarian cancer";  
RL Electrophoresis 18:686-689(1997).  
CC [1-] FUNCTION: Potent plasminogen activator and is clinically used for  
CC therapy of thrombolytic disorders.  
CC [1-] CATALYTIC ACTIVITY: Specific cleavage of Arg-|-Val bond in  
CC plasminogen to form plasmin.  
CC [1-] SUBUNIT: Found in high and low molecular mass forms. Each consists  
CC of two chains, A and B. The high molecular mass form contains a  
CC long chain A. Cleavage occurs after residue 155 in the low  
CC molecular mass form to yield a short A1 chain.  
CC [1-] PHARMACEUTICAL: Available under the name Abbokinase (Abbott). Used  
CC in Pulmonary Embolism (PE) to initiate fibrinolysis.  
CC [1-] SIMILARITY: Belongs to peptidase family S1.  
CC [1-] SIMILARITY: Contains 1 EGF-like domain.  
CC [1-] SIMILARITY: Contains 1 kringle domain.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL; X02419; CA26268.1; -

```

DR EMBL: M15476; AAAG1253.1; -
DR EMBL: D00244; BAA00175.1; -
DR EMBL: D11143; BAA01919.1; -
DR EMBL: X02760; CAA26535.1; -
DR EMBL: AF377330; AAK53822.1; -
DR EMBL: BC013575; AAH13575.1; -
DR EMBL: K03226; AAC97138.1; -
DR EMBL: K02286; AAAG1252.1; -
DR EMBL: A21571; CAA01559.1; -
DR EMBL: A18397; CAA01390.1; -
DR PIR: A00931; UKHU.
DR PDB: 1KDU; 31-OCT-93.

Query Match 36.7%; Score 131; DB 1; Length 431;
Best Local Similarity 40.3%; Pred. No. 1,1e-07;
Matches 27; Conservative 7; Mismatches 25; Indels 8; Gaps 2;

QY 1 CFWDNGHLYREDQTSAPAGRLCLNMLDA-----QSLASAPVSGAGNHSYCRNPDEDP 53
Db 70 CYEGNGHLYRGKASTDTMGRCPLPWSNATVLCQTYHAHRSDALQLGLGKHYCRNPD-NR 128

QY 54 RGPWCYV 60
Db 129 RRPWCYV 135

RESULT 11
UROK_PIG STANDARD; PRT; 442 AA.
AC P04185; 1987 (Rel. 04, Created)
DT 20-MAR-1987 (Rel. 05, Last sequence update)
DT 13-AUG-1987 (Rel. 05, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Urokinase-type plasminogen activator precursor (EC 3.4.21.73) (uPA)
DE (U-plasminogen activator).
GN PLA0.
OS Sus scrofa (Pig).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
OX NCBI_TaxID=9823;
RN [1]
SEQUENCE FROM N.A.
RC TISSUE=Kidney;
RX MEDLINE=85087954; PubMed=6096832;
RA Nagamine Y., Pearson D., Altus M.S., Reich E.;
RT "cDNA and gene nucleotide sequence of porcine plasminogen activator.";
RL Nucleic Acids Res. 12:9525-9541(1984).
RN [2]
REVISION TO 241.
RP Nagamine Y.;
RL Submitted (DEC-1986) to the PIR data bank.
CC -1- CATALYTIC ACTIVITY: Specific cleavage of Arg-|-val bond in
CC plasminogen to form plasmin.
CC -1- SIMILARITY: Belongs to peptidase family S1.
CC -1- SIMILARITY: Contains 1 EGF-like domain.
CC -1- SIMILARITY: Contains 1 kringle domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sb-sib.ch).
CC -----
DR EMBL: X01648; CAA25806.1; -
DR EMBL: X02724; CAA26511.1; -
DR PIR: A00932; UKFG.
DR HSPG: P00749; 1KDU.
DR MEROPS: S01.231; -.
DR InterPro: IPR009003; Cys Ser trypsin.
DR InterPro: IPR006209; EGF-like.
DR InterPro: IPR000001; Kringle.

```

```

DR InterPro: IPR008293; Pept_SIA_uPA.
DR InterPro: IPR001254; Peptidase_S1.
DR InterPro: IPR001314; Peptidase_S1.
DR Pfam: PF00051; kringle_1.
DR PIRSF: PIRSF001144; Urokinase_act; 1.
DR PRINTS: PR00722; CHYMOTRYPSIN.
DR PROSITE: PS00018; KRINGLE.
DR PRODOM: PD000395; Kringle; 1.
DR SMART: SM00130; KR; 1.
DR SMART: SM00020; Tryp_SPC; 1.
DR PROSITE: PS00022; EGF_1; 1.
DR PROSITE: PS01186; EGF_2; FALSE_NEG.
DR PROSITE: PS00026; EGF_3; 1.
DR PROSITE: PS00021; KRINGLE_1; 1.
DR PROSITE: PS00070; KRINGLE_2; 1.
DR PROSITE: PS00240; TRYPSIN_DOM; 1.
DR PROSITE: PS00134; TRYPSIN_HIS; 1.
DR PROSITE: PS00135; TRYPSIN_SER; 1.
DR Plasminogen activation; Hydrolase; Serine protease; Glycoprotein;
KW Kringle; EGF-like domain; Zymogen; Signal.
FT SIGNAL 1 20
FT CHAIN 21 442 UROKINASE-TYPE PLASMINOGEN ACTIVATOR.
FT CHAIN 21 188 CHAIN A (BY SIMILARITY).
FT CHAIN 190 442 CHAIN B (BY SIMILARITY).
FT DOMAIN 29 65 EGF-LIKE.
FT DOMAIN 72 153 KRINGLE.
FT DOMAIN 154 189 CONNECTING PEPTIDE.
FT DOMAIN 190 442 SERINE PROTEASE.
FT CARBOHYD 152 152 N-LINKED (GLCNAC. . .).
FT DISULFID 33 41 BY SIMILARITY.
FT DISULFID 35 53 BY SIMILARITY.
FT DISULFID 55 64 BY SIMILARITY.
FT DISULFID 179 310 INTERCHAIN (BY SIMILARITY).
FT DISULFID 220 236 BY SIMILARITY.
FT DISULFID 228 299 BY SIMILARITY.
FT DISULFID 324 393 BY SIMILARITY.
FT DISULFID 356 372 BY SIMILARITY.
FT DISULFID 383 411 BY SIMILARITY.
FT ACT_SITE 235 235 CHARGE RELAY SYSTEM.
FT ACT_SITE 286 286 CHARGE RELAY SYSTEM.
FT ACT_SITE 387 387 CHARGE RELAY SYSTEM.
FT CONFLICT 241 241 Q -> H (IN REF. 1; CAA25806).
FT CONFLICT 242 242 Q -> H (IN REF. 1; CAA26511).
FT CONFLICT 288 288 A -> GS (IN REF. 1; CAA25806).
SQ SEQUENCE 442 AA; 49116 MW; EB32FCFP50132IE CRC64;
Query Match 36.4%; Score 130; DB 1; Length 442;
Best Local Similarity 44.8%; Pred. No. 1.5e-07;
Matches 30; Conservative 4; Mismatches 25; Indels 8; Gaps 3;

QY 1 CFWDNGHLYREDQTSAPAGRLCLNMLDAQSGQL----ASAPVS---GAGNHSYCRNPDEDP 53
Db 72 CFEGNGHSYRKANTNTGRCPLPWSNATVLLNTYHAHRPDALQLGLGKHYCRNPD-NQ 130

QY 54 RGPWCYV 60
Db 131 RRPWCYV 137

RESULT 12
UROK_BOVIN STANDARD; PRT; 433 AA.
AC Q05589; Q28209;
DT 01-FEB-1994 (Rel. 28, Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Urokinase-type plasminogen activator precursor (EC 3.4.21.73) (uPA)
DE (U-plasminogen activator).
GN PLA0.
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;

```

OC Bovidae; Bovinae; Bos.  
OX NCBI\_TaxID=9913;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=ortic endothelium;  
RX MEDLINE=93216119; PubMed=8385052;  
RA Kraetzschmar J., Haendler B., Kojima S., Rifkin D.B.,  
RA Schleuning W.-D.;  
RT "Bovine urokinase-type plasminogen activator and its receptor:  
RT cloning and induction by retinoic acid.";  
RL Gene 125:177-183(1993).  
RN [2]  
RP SEQUENCE OF 12-433 FROM N.A.  
RC TISSUE=Kidney;  
RA Ravn P., Berglund L., Petersen T.E.;  
RA "Cloning and characterization of the bovine plasminogen activators uPA  
RT and tPA.";  
RL Int. Dairy J. 5:605-617(1995).  
CC -!- CATALYTIC ACTIVITY: Specific cleavage of Arg-|-Val bond in  
CC plasminogen to form plasmin.  
CC -!- INDUCTION: By retinoic acid.  
CC -!- SIMILARITY: Belongs to peptidase family S1.  
CC -!- SIMILARITY: Contains 1 EGF-like domain.  
CC -!- SIMILARITY: Contains 1 kringle domain.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; L03546; AAA51419.1; -;  
DR EMBL; X85801; CAA59796.1; -;  
DR PAR; JN0560; JN0560.  
DR HSP; P00749; ILMW.  
DR MEROPS; S01.231; -;  
DR InterPro; IPR009003; Cys Ser trypsin.  
DR InterPro; IPR006209; EGF-like.  
DR InterPro; IPR000001; Kringle.  
DR InterPro; IPR008293; Pept\_S1A\_UPA.  
DR InterPro; IPR001254; Peptidase\_S1.  
DR InterPro; IPR001314; Peptidase\_S1A.  
DR Pfam; PF00051; kringle; 1.  
DR Pfam; PF00089; trypsin; 1.  
DR PRSF; PRSF001144; Urk plas act; 1.  
DR PRINTS; PR00722; CHYMOTRYPSIN.  
DR PRINTS; PR00018; KRINGLE.  
DR ProDom; PD000395; Kringle; 1.  
DR SMART; SM00130; KR; 1.  
DR SMART; SM00020; Tryp\_Spc; 1.  
DR PROSITE; PS00022; EGF\_1; 1.  
DR PROSITE; PS01186; EGF\_2; FALSE\_NEG.  
DR PROSITE; PS50026; EGF\_3; 1.  
DR PROSITE; PS00021; KRINGLE\_1; 1.  
DR PROSITE; PS50070; KRINGLE\_2; 1.  
DR PROSITE; PS00240; TRYPSIN\_DOM; 1.  
DR PROSITE; PS00134; TRYPSIN\_HIS; 1.  
DR PROSITE; PS00135; TRYPSIN\_SER; 1.  
DR Plasminogen activation; Hydrolase; Serine protease; Glycoprotein;  
KW Kringle; EGF-like domain; Signal; Zymogen.  
FT SIGNAL 1 20  
FT CHAIN 21 433 UROKINASE-TYPE PLASMINOGEN ACTIVATOR.  
FT CHAIN 21 179 CHAIN A (BY SIMILARITY).  
FT CHAIN 181 433 CHAIN B (BY SIMILARITY).  
FT DOMAIN 29 65 EGF-LIKE.  
FT DOMAIN 72 153 KRINGLE.  
FT DOMAIN 154 180 CONNECTING PEPTIDE.  
FT DOMAIN 181 433 SERINE PROTEASE.  
FT DISULFID 33 41 BY SIMILARITY.  
FT DISULFID 53 53 BY SIMILARITY.  
FT DISULFID 55 64 BY SIMILARITY.

FT DISULFID 170 301 INTERCHAIN (BY SIMILARITY).  
FT DISULFID 211 227 BY SIMILARITY.  
FT DISULFID 219 290 BY SIMILARITY.  
FT DISULFID 315 384 BY SIMILARITY.  
FT DISULFID 347 363 BY SIMILARITY.  
FT DISULFID 374 402 BY SIMILARITY.  
FT ACT\_SITE 226 226 CHARGE RELAY SYSTEM (BY SIMILARITY).  
FT ACT\_SITE 277 277 CHARGE RELAY SYSTEM (BY SIMILARITY).  
FT ACT\_SITE 378 378 CHARGE RELAY SYSTEM (BY SIMILARITY).  
FT CONFLICT 189 189 A -> T (IN REF. 2).  
SQ SEQUENCE 433 AA; 48730 MW; 4DE1B8D4DA47027A CRC64;

Query Match 35.6%; Score 127; DB 1; Length 433;  
Best Local Similarity 41.2%; Pred.No. 3.2e-07;  
Matches 28; Conservative 6; Mismatches 24; Indels 10; Gaps 3;  
OY 1 CFWDNGHLYREDTQSPAPGLRCLNWLDAQSL-----ASAPVSGAGNHSTCRNPDED 52  
Dy 72 CYQNGHSHYRGKANRDLGSRPCLAW-DSPTVLLKMYHAHRSDAIQLGLGKHNYCRNPD-N 129  
OY 53 PRGFWCYV 60  
Dy 130 QRRPWCYV 137

RESULT 13  
IDT\_URGT\_DESRO STANDARD; PRT; 394 AA.  
AC P49150;  
DT 01-FEB-1996 (Rel. 33, Created)  
DT 01-FEB-1996 (Rel. 33, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE Salivary plasminogen activator gamma precursor (EC 3.4.21.68) (DSPA  
gamma).  
DE Desmodus rotundus (Vampire bat).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Chiroptera; Microchiroptera; Phyllostomidae;  
OC Desmodontinae; Desmodus.  
OX NCBI\_TaxID=9430;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Salivary gland;  
RX MEDLINE=92039036; PubMed=1937019;  
RA Kraetzschmar J., Haendler B., Langer G., Boidol W., Bringmann P.,  
RA Alagon A., Donner P., Schleuning W.D.;  
RT "The plasminogen activator family from the salivary gland of the  
RT vampire bat Desmodus rotundus: cloning and expression.";  
RL Gene 105:229-237(1991).  
RN [2]  
RP CHARACTERIZATION.  
RX MEDLINE=93393059; PubMed=1309059;  
RA Schleuning W.-D., Alagon A., Boidol W., Bringmann P., Petri T.,  
RA Kraetzschmar J., Haendler B., Langer G., Baldus B., Witt W.,  
RA Donner P.;  
RT "Plasminogen activators from the saliva of Desmodus rotundus (common  
RT vampire bat): unique fibrin specificity.";  
RL Ann. N.Y. Acad. Sci. 667:395-403(1992).  
CC -!- FUNCTION: Probably essential to support the feeding habits of this  
CC exclusively haematophagous animal. Probable potent thrombolytic  
CC agent.  
CC -!- CATALYTIC ACTIVITY: Specific cleavage of Arg-|-Val bond in  
CC plasminogen to form plasmin.  
CC -!- SUBUNIT: Monomer.  
CC -!- SIMILARITY: Belongs to peptidase family S1.  
CC -!- SIMILARITY: Contains 1 kringle domain.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----

```

CC -----
DR EMBL; M63990; AAA1595.1; -
DR FIR; JS0600; JS0600.
DR HSP; P98119; 1A51.
DR MEROPS; S01.239; -.
DR InterPro; IPR009003; Cys_Ser_trypsin.
DR InterPro; IPR000001; Kringle.
DR InterPro; IPR001254; Peptidase_S1.
DR Pfam; PF00051; Kringle; 1.
DR Pfam; PF00089; trypsin; 1.
DR PRINTS; PR00722; CHYMOTRYPSIN.
DR PRINTS; PR00018; KRINGLE.
DR ProDom; PD000395; Kringle; 1.
DR SMART; SM00130; KR; 1.
DR SMART; SM00020; Tryp_Spc; 1.
DR PROSITE; PS00021; KRINGLE_1; 1.
DR PROSITE; PS00070; KRINGLE_2; 1.
DR PROSITE; PS00240; TRYPSIN_DOM; 1.
DR PROSITE; PS00134; TRYPSIN_HIS; 1.
DR PROSITE; PS00135; TRYPSIN_SER; 1.
KW Plasminogen activation; Hydrolyase; Serine protease; Glycoprotein;
KW Kringle; Signal; Multigene family.
FT SIGNAL 1 36 POTENTIAL.
FT CHAIN 37 394 SALIVARY PLASMINOGEN ACTIVATOR GAMMA.
FT DOMAIN 45 126 KRINGLE.
FT ACT_SITE 142 394 SERINE PROTEASE.
FT ACT_SITE 189 189 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 238 238 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 345 345 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT DISULFID 45 126 BY SIMILARITY.
FT DISULFID 66 108 BY SIMILARITY.
FT DISULFID 97 121 BY SIMILARITY.
FT DISULFID 131 262 BY SIMILARITY.
FT DISULFID 174 190 BY SIMILARITY.
FT DISULFID 182 251 BY SIMILARITY.
FT DISULFID 276 351 BY SIMILARITY.
FT DISULFID 308 324 BY SIMILARITY.
FT DISULFID 341 369 BY SIMILARITY.
FT CARBOHYD 315 315 N-LINKED (GLCNAC...) (POTENTIAL).
SQ SEQUENCE 394 AA; 44105 MW; 9CDBF52F3D81FCD CRC64;

Query Match 35.3%; Score 126; DB 1; Length 394;
Best Local Similarity 37.3%; Pred. No. 3.8e-07;
Matches 25; Conservative 8; Mismatches 26; Indels 8; Gaps 2;

QY 1 CFWDNGLHYREDTSPAPGRLNW-----LDAOSGLASAPVSGAGNHSYCRNPDEDP 53
   | : ||| : : : : : : : : : : : : : : : : : : : : : : : :
Db 45 CYKQGVYRGVTSTSESGACINWNSNLLIRTYNGRYPEAVKLGHNHYCRNPDGAS 104
   | : ||| : : : : : : : : : : : : : : : : : : : : : : : :

QY 54 RGPWCYV 60
   | : ||| : : : : : : : : : : : : : : : : : : : : : : : :
Db 105 K-PWCYV 110

RESULT 14
HGFL_MOUSE STANDARD; PRT; 716 AA.
AC P26928;
DT 01-AUG-1992 (Rel. 23, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Hepatocyte growth factor-like protein precursor (Macrophage
DE stimulatory protein) (MSP).
GN MST1 OR HGFL
OS Mus musculus (Mouse);
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BALB/c; TISSUE=Liver;
RM MEDLINE=92002017; PubMed=1832957;

```

```

RA Friezner Degen S.J., Stuart L.A., Han S., Jamison C.S.;
RT "Characterization of the mouse cDNA and gene coding for a hepatocyte
RL growth factor-like protein: expression during development.";
CC Biochemistry 30:9781-9791(1991).
CC -!- FUNCTION: Probably has no proteolytic activity, since crucial AA
CC characteristic of serine proteases catalytic sites are not
CC conserved.
CC -!- TISSUE SPECIFICITY: Liver. Lower levels in lung, placenta and
CC adrenal.
CC -!- DEVELOPMENTAL STAGE: Is expressed at low levels during gestation.
CC just before birth the level increases dramatically and remains
CC stable afterwards.
CC -!- PTM: MAY BE CLEAVED AFTER AA 488, TO YIELD A TWO-CHAIN MOLECULE
CC HELD TOGETHER BY DISULFIDE BONDS, OR CLEAVED INTO TWO SEPARATE
CC POLYPEPTIDES.
CC -!- SIMILARITY: Belongs to peptidase family S1. Plasminogen subfamily.
CC -!- SIMILARITY: Contains 4 kringle domains.
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; M74180; AAA50166.1; -.
CC EMBL; M74181; AAA50167.1; -.
CC PIR; A40332; A40332.
CC HSP; P00747; 1KRN.
CC MEROPS; S01.975; -.
CC MGD; MGI:96080; Mst1.
CC GO; GO:0007586; P:embryo implantation; IC.
CC InterPro; IPR009003; Cys_Ser_trypsin.
CC InterPro; IPR000001; Kringle.
CC InterPro; IPR003014; PAN.
CC InterPro; IPR003609; Pan_app.
CC InterPro; IPR001254; Peptidase_S1.
CC InterPro; IPR001314; Peptidase_S1A.
CC InterPro; IPR003966; Peptidase_S1A_pr.
CC Pfam; PF00051; Kringle; 4.
CC Pfam; PF00024; PAN; 1.
CC Pfam; PF00089; trypsin; 1.
CC PRINTS; PR00722; CHYMOTRYPSIN.
CC PRINTS; PR00018; KRINGLE.
CC PRINTS; PR01505; BROTHROMBIN.
CC ProDom; PD000395; Kringle; 4.
CC SMART; SM00130; KR; 4.
CC SMART; SM00473; PAN_AP; 1.
CC SMART; SM00020; Tryp_Spc; 1.
CC PROSITE; PS00021; KRINGLE_1; 4.
CC PROSITE; PS00070; KRINGLE_2; 4.
CC PROSITE; PS00240; TRYPSIN_DOM; 1.
KW Kringle; Glycoprotein; Serine protease homolog; Repeat; Signal.
FT SIGNAL 1 18 POTENTIAL.
FT CHAIN 19 716 HEPATOCYTE GROWTH FACTOR-LIKE PROTEIN.
FT DOMAIN 19 109 PAP.
FT DOMAIN 110 186 KRINGLE 1.
FT DOMAIN 191 268 KRINGLE 2.
FT DOMAIN 292 370 KRINGLE 3.
FT DOMAIN 379 457 KRINGLE 4.
FT DOMAIN 489 716 SERINE PROTEASE-LIKE.
FT DISULFID 56 78 BY SIMILARITY.
FT DISULFID 60 66 BY SIMILARITY.
FT DISULFID 110 186 BY SIMILARITY.
FT DISULFID 131 169 BY SIMILARITY.
FT DISULFID 157 181 BY SIMILARITY.
FT DISULFID 191 268 BY SIMILARITY.
FT DISULFID 194 333 INTERCHAIN (BY SIMILARITY).
FT DISULFID 212 251 BY SIMILARITY.
FT DISULFID 240 263 BY SIMILARITY.
FT DISULFID 292 370 BY SIMILARITY.
FT DISULFID 313 352 BY SIMILARITY.

```

FT DISULFID 341 364 BY SIMILARITY.  
FT DISULFID 379 457 BY SIMILARITY.  
FT DISULFID 400 440 BY SIMILARITY.  
FT DISULFID 428 452 BY SIMILARITY.  
FT DISULFID 477 593 INTERCHAIN (BY SIMILARITY).  
FT DISULFID 512 528 BY SIMILARITY.  
FT DISULFID 607 672 BY SIMILARITY.  
FT DISULFID 637 651 BY SIMILARITY.  
FT DISULFID 662 690 BY SIMILARITY.  
FT CARBOHYD 72 72 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 173 173 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 305 305 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 620 620 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CONFLICT 19 19 P -> Q (IN GENOMIC SEQUENCE).  
SQ SEQUENCE 716 AA; 80588 MW; BBCE02EF95213ACC CRC64;  
  
Query Match 35.3%; Score 126; DB 1; Length 716;  
Best Local Similarity 41.9%; Pred. No. 7e-07;  
Matches 26; Conservative 3; Mismatches 29; Indels 4; Gaps 2;  
  
QY 1 CFWDNGHLYREDQTSAPGLRCLNW---LDAQSLASAPVSGAGNHSCRNPDDEPRGPW 57  
DB 110 CIMDNGSVYRGTVARTAGLPCQAWSRFPNDHKYTPTRKGL-EENFCRNPDDGPRGPW 168  
  
QY 58 CY 59  
DB 169 CY 170  
  
RESULT 15  
PLMN RAT STANDARD; PRT; 169 AA.  
AC Q01177;  
DT 01-APR-1993 (Rel. 25, Created)  
DT 01-APR-1993 (Rel. 25, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE Plasminogen (EC 3.4.21.7) (Fragment).  
GN PLG.  
OS Rattus norvegicus (Rat).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
OX NCBI\_TaxID=10116;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Liver;  
RX MEDLINE=91250378; PubMed=1645711;  
RA Kanalas J.J., Makker S.P.  
RT "Identification of the rat Heymann nephritis autoantigen (GP330) as a receptor site for plasminogen.";  
RL J. Biol. Chem. 266:10825-10829(1991).  
CC -!- FUNCTION: Plasmin dissolves the fibrin of blood clots and acts as a proteolytic factor in a variety of other processes including embryonic development, tissue remodeling, tumor invasion, and inflammation; in ovulation it weakens the walls of the Graafian follicle. It activates the urokinase-type plasminogen activator, collagenases and several complement zymogens, such as C1 and C5. It cleaves fibrin, fibronectin, thrombospondin, laminin and von Willebrand factor.  
CC -!- CATALYTIC ACTIVITY: Preferential cleavage: Lys-|-Xaa > Arg-|-Xaa; higher selectivity than trypsin. Converts fibrin into soluble products.  
CC -!- ENZYME REGULATION: Converted into plasmin by plasminogen activators, both plasminogen and its activator being bound to fibrin. Cannot be activated with streptokinase.  
CC -!- MISCELLANEOUS: Plasmin is inactivated by alpha-2-antiplasmin immediately after dissociation from the clot.  
CC -!- SIMILARITY: Belongs to peptidase family S1. Plasminogen subfamily.  
CC -!- SIMILARITY: Contains 5 kringle domains.  
-----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----

CC EMBL; M62832; AAA41884.1; -  
DR PIR; A40522; A40522.  
DR HSP; P00747; IPMK.  
DR MEROPS; S01.233; -  
DR InterPro; IPR00001; Kringle.  
DR InterPro; IPR001254; Peptidase\_S1.  
DR InterPro; IPR003966; Peptidase\_S1A\_pr.  
DR Pfam; PF00051; kringle; 2.  
DR PRINTS; PR0018; KRINGLE  
DR PRINTS; PR01505; PROTHROMBIN.  
DR ProDom; PD000395; Kringle; 2.  
DR SMART; SMC0130; KR; 1.  
DR PROSITE; PS00021; KRINGLE 1; 1.  
DR PROSITE; PS50070; KRINGLE 2; 2.  
DR PROSITE; PS0240; TRYPSIN\_DOM; PARTIAL.  
DR PROSITE; PS0134; TRYPSIN\_HIS; PARTIAL.  
DR PROSITE; PS00135; TRYPSIN\_SER; PARTIAL.  
KW Hydrolase; Serine protease; Plasma; Glycoprotein; Fibrinolysis;  
KW Tissue remodeling; Blood coagulation; Kringle; Repeat.  
FT NON\_TER 1 1  
FT DOMAIN <1 10 KRINGLE 3 (BY SIMILARITY).  
FT DOMAIN 34 112 KRINGLE 4 (BY SIMILARITY).  
FT DOMAIN 139 >169 KRINGLE 5 (BY SIMILARITY).  
FT DISULFID 34 112 BY SIMILARITY.  
FT DISULFID 55 95 BY SIMILARITY.  
FT DISULFID 83 107 BY SIMILARITY.  
FT NON\_TER 169 169  
SQ SEQUENCE 169 AA; 18401 MW; 77A54214C49D010C CRC64;

Query Match 34.5%; Score 123; DB 1; Length 169;  
Best Local Similarity 37.5%; Pred. No. 3.4e-07;  
Matches 24; Conservative 9; Mismatches 25; Indels 6; Gaps 2;

QY 1 CFWDNGHLYREDQTSAPGLRCLNWLD-----QSLASAPVSGAGNHSCRNPDDEPRG 55  
DB 34 CYQNGKSVYRGTSSTNTGKQSVMTSPHSHSKTPANFPDGL-EMNYCRNPNDQRG 92  
  
QY 56 PWCV 59  
DB 93 PWCF 96

Search completed: March 17, 2004, 07:04:25  
Job time : 2.42261 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 17, 2004, 06:59:03 ; Search time 10.9825 Seconds  
(without alignments)  
1723.750 Million cell updates/sec

Title: US-10-057-951-2\_COPY\_25\_84

Perfect score: 357

Sequence: 1 CFWDNGHLYREDQTSFAPGL.....GNHSYCRNPDPDRGPWCYV 60

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL\_25.\*

- 1: sp\_archaea.\*
- 2: sp\_bacteria.\*
- 3: sp\_fungi.\*
- 4: sp\_human.\*
- 5: sp\_invertebrate.\*
- 6: sp\_mammal.\*
- 7: sp\_mhc.\*
- 8: sp\_organelle.\*
- 9: sp\_phase.\*
- 10: sp\_plant.\*
- 11: sp\_rodent.\*
- 12: sp\_virus.\*
- 13: sp\_vertebrate.\*
- 14: sp\_unclassified.\*
- 15: sp\_rvirus.\*
- 16: sp\_bacteriaph.\*
- 17: sp\_archaeap.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query % Match	Length	ID	Description
1	357	100.0	234	4 Q86YW2	Q86YW2 homo sapien
2	357	100.0	263	4 O00318	O00318 homo sapien
3	357	100.0	263	4 Q8NCJ9	Q8NCJ9 homo sapien
4	357	100.0	263	4 Q96FE7	Q96FE7 homo sapien
5	280	78.4	213	11 Q81123	Q81123 mus musculus
6	280	78.4	264	11 Q81122	Q81122 mus musculus
7	280	78.4	264	11 Q7TMJ8	Q7TMJ8 mus musculus
8	155	43.4	562	6 Q8SQ23	Q8SQ23 sus scrofa
9	146	40.9	291	4 Q7Z7N2	Q7Z7N2 homo sapien
10	146	40.9	516	4 Q9BU99	Q9BU99 homo sapien
11	146	40.9	562	4 Q86YK8	Q86YK8 homo sapien
12	146	40.9	653	11 Q8VCS4	Q8VCS4 mus musculus
13	143	40.1	564	6 Q8MKB1	Q8MKB1 oryctolagus
14	131	36.7	154	4 Q96SE8	Q96SE8 homo sapien
15	130.5	36.6	709	13 Q7ZTN9	Q7ZTN9 xenopus lae
16	130.5	36.6	717	13 P70006	P70006 xenopus lae

17	129.5	36.3	421	13 Q8AXX3	Q8AXX3 xenopus lae
18	128	35.9	90	4 Q8NG20	Q8NG20 homo sapien
19	128	35.9	716	11 P70521	P70521 rattus norv
20	127	35.6	157	6 Q9TVAB	Q9TVAB bos taurus
21	127	35.6	313	13 Q9PU78	Q9PU78 crocodylus
22	127	35.6	395	4 Q9BZW1	Q9BZW1 homo sapien
23	126	35.3	716	11 Q91XG8	Q91XG8 mus musculu
24	124	34.7	263	13 Q7SXB3	Q7SXB3 brachydanio
25	122.5	34.3	716	13 Q91691	Q91691 xenopus lae
26	122	34.2	704	13 Q90865	Q90865 gallus gall
27	121	33.9	810	4 Q15146	Q15146 homo sapien
28	120	33.6	812	11 Q9R0W3	Q9R0W3 rattus norv
29	119	33.3	231	11 Q8C6L2	Q8C6L2 mus musculu
30	118	33.1	728	6 Q9BH09	Q9BH09 felis silve
31	118	33.1	728	11 Q8C9G5	Q8C9G5 mus musculu
32	118	33.1	730	6 Q867B7	Q867B7 canis fami
33	118	33.1	806	6 O18783	O18783 macropus eu
34	115.5	32.4	391	5 Q86FQ9	Q86FQ9 cryptospori
35	115	32.2	433	6 Q8MIL0	Q8MIL0 oryctolagus
36	115	32.2	433	6 Q8MHY7	Q8MHY7 oryctolagus
37	115	32.2	540	13 Q800Y7	Q800Y7 meleagris g
38	113.5	31.8	95	4 Q8N696	Q8N696 homo sapien
39	112.5	31.5	334	6 Q46507	Q46507 papio hamad
40	111	31.1	385	5 Q25101	Q25101 herdmania m
41	109.5	30.7	759	11 Q7TF84	Q7TF84 rattus norv
42	109	30.5	202	13 Q90675	Q90675 gallus gall
43	109	30.5	567	4 Q13208	Q13208 homo sapien
44	108.5	30.4	616	6 Q97507	Q97507 sus scrofa
45	108	30.3	215	13 O42341	O42341 gallus gall

#### ALIGNMENTS

RESULT 1

Q86YW2 PRELIMINARY; PRT; 234 AA.  
 ID Q86YW2 PRELIMINARY; PRT; 234 AA.  
 AC Q86YW2; 2003 (T-EMBLrel. 24, Created)  
 DT 01-JUN-2003 (T-EMBLrel. 24, Last sequence update)  
 DT 01-OCT-2003 (T-EMBLrel. 25, Last annotation update)  
 DE HGFL(S) protein.  
 GN HGFL.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Chiang H., Chang M.;  
 RL Submitted (JUL-2002) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AF528079; AAC33762.1; .  
 DR InterPro; IPR000001; Kringle.  
 DR Pfam; PF00051; Kringle; 1.  
 DR PRINTS; PRO0018; KRINGLE.  
 DR SMART; SMO0130; KR; 1.  
 DR PROSITE; PS00021; KRINGLE\_1; 1.  
 DR PROSITE; PS00070; KRINGLE\_2; 1.  
 SQ SEQUENCE 234 AA; 25320 MW; C78F64BDBD1B8DCD CRC64;

Query Match 100.0%; Score 357; DB 4; Length 234;

Best Local Similarity 100.0%; Pred. No. 1.3e-34;

Matches 60; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CFWDNGHLYREDQTSFAPGLRCINLWLDQAQSLASAPVSGAGNHSYCRNPDPDRGPWCYV 60

Db 25 CFWDNGHLYREDQTSFAPGLRCINLWLDQAQSLASAPVSGAGNHSYCRNPDPDRGPWCYV 84

RESULT 2

O00318 PRELIMINARY; PRT; 263 AA.  
 ID O00318  
 AC O00318;

01-JUL-1997 (TrEMBLrel. 04, Created)  
 01-JUL-1997 (TrEMBLrel. 04, Last sequence update)  
 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
 WUGSC:DJ515N1.2 protein.  
 GN WUGSC:DJ515N1.2  
 OS Homo sapiens (Human)  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RA DU Z., Scheet P., Harper M.;  
 RP "The sequence of H. sapiens PAC clone RP3-515N1.";  
 RL Submitted (MAY-1997) to the EMBL/GenBank/DBJ databases.  
 RV [2]  
 RP SEQUENCE FROM N.A.  
 RA Waterston R.;  
 RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.  
 CC -1- SIMILARITY: CONTAINS 1 KRINGLE DOMAIN.  
 DR EMBL; AC002073; AAB54054.1; -;  
 DR HSSP; P00749; 1KDU  
 DR InterPro; IPR000001; Kringle.  
 DR PRINTS; PR00051; Kringle; 1.  
 DR PRODOM; PD000395; Kringle; 1.  
 DR SMART; SM00130; KR; 1.  
 DR PROSITE; PS00021; KRINGLE\_1; 1.  
 DR PROSITE; PS00070; KRINGLE\_2; 1.  
 KW Glycoprotein; Kringle  
 SQ SEQUENCE 263 AA; 28248 MW; 197C3EE8B54A242 CRC64;  
 Query Match 100.0%; Score 357; DB 4; Length 263;  
 Best Local Similarity 100.0%; Pred. No. 1.5e-34;  
 Matches 60; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 CFWDNGHLYREDQTSAPGLRCLNWLDAQSLASAPVSGAGNHSYCRNPDEDPGPGWCYV 60  
 DB 25 CFWDNGHLYREDQTSAPGLRCLNWLDAQSLASAPVSGAGNHSYCRNPDEDPGPGWCYV 84  
 RESULT 3  
 Q8NCJ9 PRELIMINARY; PRT; 263 AA.  
 ID Q8NCJ9  
 AC Q8NCJ9  
 DT 01-OCT-2002 (TrEMBLrel. 22, Created)  
 DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)  
 DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)  
 DE Hypothetical protein FLJ90207.  
 OS Homo sapiens (Human)  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RA Hattori A., Okumura K., Iwayanagi T., Ninomiya K.;  
 RP "NEDO human cDNA sequencing project.";  
 RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.  
 CC -1- SIMILARITY: CONTAINS 1 KRINGLE DOMAIN.  
 DR EMBL; AK074688; BAC11140.1; -;  
 DR InterPro; IPR000001; Kringle.  
 DR Pfam; PF00051; Kringle; 1.  
 DR PRINTS; PR00018; KRINGLE.  
 DR PRODOM; PD000395; Kringle; 1.  
 DR SMART; SM00130; KR; 1.  
 DR PROSITE; PS00021; KRINGLE\_1; 1.  
 DR PROSITE; PS00070; KRINGLE\_2; 1.  
 KW Hypothetical protein; Glycoprotein; Kringle.  
 SQ SEQUENCE 263 AA; 28104 MW; 73A9294D5142C6C3 CRC64;

Query Match 100.0%; Score 357; DB 4; Length 263;  
 Best Local Similarity 100.0%; Pred. No. 1.5e-34;  
 Matches 60; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 CFWDNGHLYREDQTSAPGLRCLNWLDAQSLASAPVSGAGNHSYCRNPDEDPGPGWCYV 60  
 DB 25 CFWDNGHLYREDQTSAPGLRCLNWLDAQSLASAPVSGAGNHSYCRNPDEDPGPGWCYV 84  
 RESULT 4  
 Q96FE7 PRELIMINARY; PRT; 263 AA.  
 ID Q96FE7  
 AC Q96FE7  
 DT 01-DEC-2001 (TrEMBLrel. 19, Created)  
 DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)  
 DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)  
 DE Hypothetical protein (HGFL(L) protein).  
 GN HGFL.  
 OS Homo sapiens (Human)  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RA SEQUENCE FROM N.A.  
 RP TISSUE=Brain;  
 RC Strausberg R.;  
 RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.  
 RV [2]  
 RP SEQUENCE FROM N.A.  
 RA Chiang H., Chang M.;  
 RL Submitted (JUL-2002) to the EMBL/GenBank/DBJ databases.  
 CC -1- SIMILARITY: CONTAINS 1 KRINGLE DOMAIN.  
 DR EMBL; BC011049; AAH1049.1; -;  
 DR EMBL; AF528080; AAO33763.1; -;  
 DR InterPro; IPR000001; Kringle.  
 DR Pfam; PF00051; Kringle; 1.  
 DR PRINTS; PR00018; KRINGLE.  
 DR PRODOM; PD000395; Kringle; 1.  
 DR SMART; SM00130; KR; 1.  
 DR PROSITE; PS00021; KRINGLE\_1; 1.  
 DR PROSITE; PS00070; KRINGLE\_2; 1.  
 KW Hypothetical protein; Glycoprotein; Kringle.  
 SQ SEQUENCE 263 AA; 28234 MW; 197C3EE8B88FA242 CRC64;  
 Query Match 100.0%; Score 357; DB 4; Length 263;  
 Best Local Similarity 100.0%; Pred. No. 1.5e-34;  
 Matches 60; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 CFWDNGHLYREDQTSAPGLRCLNWLDAQSLASAPVSGAGNHSYCRNPDEDPGPGWCYV 60  
 DB 25 CFWDNGHLYREDQTSAPGLRCLNWLDAQSLASAPVSGAGNHSYCRNPDEDPGPGWCYV 84  
 RESULT 5  
 Q811Z3 PRELIMINARY; PRT; 213 AA.  
 ID Q811Z3  
 AC Q811Z3  
 DT 01-JUN-2003 (TrEMBLrel. 24, Created)  
 DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)  
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
 DE HGFL(S) protein.  
 GN HGFL.  
 OS Mus musculus (Mouse)  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RA SEQUENCE FROM N.A.  
 RP STRAIN=BALE/c;  
 RC Chiang H., Chang M.;  
 RL Submitted (JUL-2002) to the EMBL/GenBank/DBJ databases.  
 RV [2]  
 RP SEQUENCE 263 AA; 28334 MW; 197C3EE8B88FA242 CRC64;



RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,  
RA Stapleton M., Soares M.B., Donald M.F., Casavant T.L., Scheetz T.B.,  
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,  
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,  
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,  
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
RA Fahey J., Helton E., Kettman M., Madan A., Rodrigues S., Sanchez A.,  
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
RA Blakeley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,  
RA Krzywinski M.I., Skalska U., Smallog U., Schnerch D.E., Schein J.E.,  
RA Jones S.J., Marra M.A.;  
RT "Generation and initial analysis of more than 15,000 full-length human  
RT and mouse cDNA sequences.";  
RT Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).  
RN [2]  
RN SEQUENCE FROM N.A.  
RC STRAIN=C57BL/6J; TISSUE=Breast tumor;  
RL Straussberg R.;  
RL Submitted (AUG-2003) to the EMBL/GenBank/DBSJ databases.  
RL EMBL; BC055920; AAH55920.1; -.  
KW Hypothetical protein.  
SQ SEQUENCE 264 AA; 28567 MW; 833EA578FEB763A4 CRC64;  
  
Query Match 78.4%; Score 280; DB 11; Length 264;  
Best Local Similarity 75.0%; Pred. No. 2.4e-25;  
Matches 45; Conservative 3; Mismatches 12; Indels 0; Gaps  
  
Qy 1 CFWDNGHLYREDQTSPPAGRLCLNWLDAQSGILASPVSGAGNHSYCRNPDEPRGFWCYV 60  
Db 25 CFWDNGHLYREDQTSPPAGRLCLNWLDAQSGRESLTFSPGNHNYCRNPQDPRGFWCYI 84  
|||||  
  
RESULT 8  
Q8SQ23 PRELIMINARY; PRT; 562 AA.  
ID Q8SQ23  
AC Q8SQ23;  
DT 01-JUN-2002 (T=EMBLrel. 21, Created)  
DT 01-JUN-2002 (T=EMBLrel. 21, Last sequence update)  
DT 01-OCT-2003 (T=EMBLrel. 25, Last annotation update)  
DE T-plasminogen activator.  
DE Sus scrofa (Pig).  
OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.  
OC NCBI\_TaxId=9823;  
RN [1]  
RN SEQUENCE FROM N.A.  
RC TISSUE=Enamel organ;  
RA Ding Y., Xue J., Bartlett J.D.;  
RT "T-plasminogen activator in tooth tissues.";  
RT Submitted (MAR-2001) to the EMBL/GenBank/DBSJ databases.  
CC -1- SIMILARITY: CONTAINS 2 KRINGLE DOMAINS.  
DR EMBL; AF3646505; AAM00297.1; -.  
DR HSP; P00761; 1AN1.  
DR GO; GO:0005576; C:extracellular; IEA.  
DR GO; GO:0004263; F:chymotrypsin activity; IEA.  
DR GO; GO:0008233; F:peptidase activity; IEA.  
DR GO; GO:0004295; F:trypsin activity; IEA.  
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.  
DR InterPro; IPR009003; Cys\_Ser\_trypsin.  
DR InterPro; IPR06209; EGF\_like.  
DR InterPro; IPR000083; Fibrinctn1.  
DR InterPro; IPR06210; IEGF.  
DR InterPro; IPR000001; Kringle.  
DR InterPro; IPR001254; Peptidase S1.  
DR InterPro; IPR001314; Peptidase\_S1A.  
DR Pfam; PF00008; EGF; 1.  
DR Pfam; PF00039; fn1; 1.  
DR Pfam; PF00051; kringle; 2.  
DR Pfam; PF00089; trypsin; 1.  
DR PRINTS; PR00722; CHYMOTRYPSIN.  
DR PRINTS; PR00018; KRINGLE.

DR ProDom; PD000395; Kringle; 2.  
 DR SMART; SM00181; EGF; 1.  
 DR SMART; SM00058; FN1; 1.  
 DR SMART; SM00130; KR; 2.  
 DR SMART; SM00020; Tryp\_SPC; 1.  
 DR PROSITE; PS00022; EGF\_1; 1.  
 DR PROSITE; PS01186; EGF\_2; 1.  
 DR PROSITE; PS01253; FIBRONECTIN\_1; 1.  
 DR PROSITE; PS00021; KRINGLE\_1; 2.  
 DR PROSITE; PS00070; KRINGLE\_2; 2.  
 DR PROSITE; PS00240; TRYPSIN\_DOM; 1.  
 DR PROSITE; PS00134; TRYPSIN\_HIS; 1.  
 DR PROSITE; PS00135; TRYPSIN\_SER; 1.  
 KW EGF-like domain; Glycoprotein; Hydrolase; Kringle; Protease;  
 KW Serine protease.  
 SQ SEQUENCE 562 AA; 63668 MW; F9E6B4C77CB101E8 CRC64;

Query Match 43.4%; Score 155; DB 6; Length 562;  
 Best Local Similarity 43.5%; Pred. No. 4.7e-10;  
 Matches 30; Conservative 9; Mismatches 18; Indels 12; Gaps 3;

QY 1 CFWDNGHLYREDQTSAPGLRCLNWLDAQSLASAPVS-----GAGNHSYCRNPDE 51  
 DB 127 CYEDQGISYRGTSWTAESGAECTNW--NTSGLASMPYNGRRPDVAKVLGLGNHNYCRNPDK 184

QY 52 DPGPWCYV 60  
 DB 185 DSK-PWCYV 192

RESULT 9  
 Q7Z7N2 PRELIMINARY; PRT; 291 AA.  
 AC Q7Z7N2;  
 DT 01-OCT-2003 (TrEMBLrel. 25, Created)  
 DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)  
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
 DE Plasminogen activator, tissue type isoform 2.  
 GN PLAT.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Rieder M.J., Carrington D.P., da Ponte S.H., Hastings N.C.,  
 RA Ahearn M.O., Kuldaneck S.A., Rajkumar N., Toth E.J., Yi Q.,  
 RA Nickerson D.A.;  
 RL Submitted (MAY-2003) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AY291860; AAP34246.1; -- 874E38C52F50EF1D CRC64;  
 SQ SEQUENCE 291 AA; 32191 MW; 874E38C52F50EF1D CRC64;

Query Match 40.9%; Score 146; DB 4; Length 291;  
 Best Local Similarity 43.5%; Pred. No. 2.7e-09;  
 Matches 30; Conservative 5; Mismatches 22; Indels 12; Gaps 3;

QY 1 CFWDNGHLYREDQTSAPGLRCLNWLDAQSLASAPVS-----GAGNHSYCRNPDE 51  
 DB 127 CYEDQGISYRGTSWTAESGAECTNW--NSSALACKPYSGRRPDALRLGLGNHNYCRNPDR 184

QY 52 DPGPWCYV 60  
 DB 185 DSK-PWCYV 192

RESULT 10  
 Q9BU99 PRELIMINARY; PRT; 516 AA.  
 AC Q9BU99;  
 DT 01-JUN-2001 (TrEMBLrel. 17, Created)  
 DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)  
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
 DE Similar to plasminogen activator, tissue.

OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Skin;  
 RA Strausberg R.;  
 RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.  
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1.  
 CC -1- SIMILARITY: CONTAINS 2 KRINGLE DOMAINS.  
 DR EMBL; BC002795; AAH02795.1; --  
 DR HSP; P00750; IASH.  
 DR GO; GO:0004263; F-chymotrypsin activity; IEA.  
 DR GO; GO:0008233; F-peptidase activity; IEA.  
 DR GO; GO:0004295; F-trypsin activity; IEA.  
 DR GO; GO:0006508; P-proteolysis and peptidolysis; IEA.  
 DR InterPro; IPR009003; Cys Ser trypsin.  
 DR InterPro; IPR006209; EGF-like.  
 DR InterPro; IPR006210; IEGF.  
 DR InterPro; IPR000001; Kringle.  
 DR InterPro; IPR001254; Peptidase\_S1.  
 DR InterPro; IPR001314; Peptidase\_S1A.  
 DR Pfam; PFO0008; EGF; 1.  
 DR Pfam; PFO0051; kringle; 2.  
 DR PRINTS; PR00722; CHYMOTRYPSIN.  
 DR PRINTS; PR00018; KRINGLE.  
 DR ProDom; PD000395; Kringle; 2.  
 DR SMART; SM00181; EGF; 1.  
 DR SMART; SM00130; KR; 2.  
 DR SMART; SM00020; Tryp\_SPC; 1.  
 DR PROSITE; PS00022; EGF\_1; 1.  
 DR PROSITE; PS01186; EGF\_2; 1.  
 DR PROSITE; PS00021; KRINGLE\_1; 2.  
 DR PROSITE; PS00070; KRINGLE\_2; 2.  
 DR PROSITE; PS00240; TRYPSIN\_DOM; 1.  
 DR PROSITE; PS00134; TRYPSIN\_HIS; 1.  
 DR PROSITE; PS00135; TRYPSIN\_SER; 1.  
 KW EGF-like domain; Glycoprotein; Hydrolase; Kringle; Protease;  
 KW Serine protease.  
 SQ SEQUENCE 516 AA; 57370 MW; BAB31901FDC96800 CRC64;

Query Match 40.9%; Score 146; DB 4; Length 516;  
 Best Local Similarity 43.5%; Pred. No. 5.1e-09;  
 Matches 30; Conservative 5; Mismatches 22; Indels 12; Gaps 3;

QY 1 CFWDNGHLYREDQTSAPGLRCLNWLDAQSLASAPVS-----GAGNHSYCRNPDE 51  
 DB 81 CYEDQGISYRGTSWTAESGAECTNW--NSSALACKPYSGRRPDALRLGLGNHNYCRNPDR 138

QY 52 DPGPWCYV 60  
 DB 139 DSK-PWCYV 146

RESULT 11  
 Q86YK8 PRELIMINARY; PRT; 562 AA.  
 ID Q86YK8;  
 AC Q86YK8;  
 DT 01-JUN-2003 (TrEMBLrel. 24, Created)  
 DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)  
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
 DE Tissue plasminogen activator.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Liu Y., Xu L., Zeng Y., He X.;  
 RA "cDNA of tissue plasminogen activator."  
 RT Submitted (JAN-2003) to the EMBL/GenBank/DBJ databases.  
 RL

```

DR EMBL; AY221101; AA034406.1; -
DR GO; GO:0005576; C:extracellular; IEA.
DR GO; GO:0004263; F:chymotrypsin activity; IEA.
DR GO; GO:0004295; F:trypsin activity; IEA.
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR InterPro; IPR009003; Cys Ser trypsin.
DR InterPro; IPR006209; EGF_like.
DR InterPro; IPR000083; Fibrinctnl.
DR InterPro; IPR000001; Kringle.
DR InterPro; IPR001254; Peptidase_S1.
DR Pfam; PF00008; EGF; 1.
DR Pfam; PF00039; fn1; 1.
DR Pfam; PF00040; fn2; 1.
DR Pfam; PF00051; kringle; 1.
DR Pfam; PF00089; trypsin; 1.
DR PRINTS; PRO0722; CHYMOTRYPSIN.
DR PRINTS; PRO0018; Kringle.
DR PRINTS; PRO0083; trypsin_1.
DR ProDom; PD000395; Kringle; 2.
DR SMART; SM00058; FN1; 1.
DR SMART; SM00130; KR; 1.
DR SMART; SM00020; Tryp_SPC; 1.
DR PROSITE; PS00022; EGF_1; 1.
DR PROSITE; PS01186; EGF_2; 1.
DR PROSITE; PS01253; FIBRONECTIN_1; 1.
DR PROSITE; PS00021; KRINGLE_1; 1.
DR PROSITE; PS00070; KRINGLE_2; 2.
DR PROSITE; PS02040; TRYPSIN_DOM; 1.
DR PROSITE; PS00134; TRYPSIN_HIS; 1.
DR PROSITE; PS00135; TRYPSIN_SER; 1.
SQ SEQUENCE 562 AA; 62902 MW; 837D98392F6EDD1F CRC64;

Query Match 40.9%; Score 146; DB 4; Length 562;
Best Local Similarity 43.5%; Pred. No. 5.5e-09;
Matches 30; Conservative 5; Mismatches 22; Indels 12; Gaps 3;

QY 1 CFVNGHLYREDQTSAPGLRCNLWDAQSLASAPVS-----GAGNHSCYCRNPDE 51
Db 127 CYEQGSGYRTWTASGAECTNW--NSSALQKPYSGRRPDAIRLIGLNHNYCRNPDR 184
QY 52 DPGPGWCYV 60
Db 185 DSK-PWCYV 192

RESULT 12
Q8VCS4 PRELIMINARY; PRT; 653 AA.
AC Q8VCS4;
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Hypothetical protein.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OC NCBI_TaxId=10090;
RN [1]
RP TISSUE=Liver;
RA Strausberg R.;
RL Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1.
CC -1- SIMILARITY: CONTAINS 1 KRINGLE DOMAIN.
DR EMBL; BC019376; AAH19376.1; -.
DR HSSP; P00761; 1AN1
DR GO; GO:0005576; C:extracellular; IEA.
DR GO; GO:0004263; F:chymotrypsin activity; IEA.
DR GO; GO:0008233; F:peptidase activity; IEA.
DR GO; GO:0004295; F:trypsin activity; IEA.
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR InterPro; IPR009003; Cys Ser trypsin.
DR InterPro; IPR000742; EGF_2.
DR InterPro; IPR006209; EGF_like.

```

```

DR InterPro; IPR000083; Fibrinctnl.
DR InterPro; IPR000562; FN_Type_II.
DR InterPro; IPR006210; IEGF.
DR InterPro; IPR000001; Kringle.
DR InterPro; IPR001254; Peptidase_S1.
DR InterPro; IPR001314; Peptidase_SIA.
DR Pfam; PF00008; EGF; 2.
DR Pfam; PF00039; fn1; 1.
DR Pfam; PF00040; fn2; 1.
DR Pfam; PF00051; kringle; 1.
DR Pfam; PF00089; trypsin; 1.
DR PRINTS; PRO0722; CHYMOTRYPSIN.
DR PRINTS; PRO0018; Kringle.
DR ProDom; PD000395; FN_Type_II; 1.
DR SMART; SM00181; EGF; 2.
DR SMART; SM00059; FN2; 1.
DR SMART; SM00130; KR; 1.
DR SMART; SM00020; Tryp_SPC; 1.
DR PROSITE; PS00022; EGF_1; 2.
DR PROSITE; PS01186; EGF_2; 1.
DR PROSITE; PS01253; FIBRONECTIN_1; 1.
DR PROSITE; PS00021; KRINGLE_1; 1.
DR PROSITE; PS00070; KRINGLE_2; 1.
DR PROSITE; PS02040; TRYPSIN_DOM; 1.
DR PROSITE; PS00134; TRYPSIN_HIS; 1.
DR PROSITE; PS00135; TRYPSIN_SER; 1.
KW Hypothetical protein; EGF-like domain; Glycoprotein; Hydrolase;
KW Kringle; Protease; Serine protease.
SQ SEQUENCE 653 AA; 70553 MW; FE18D90174EB6FDD CRC64;

Query Match 40.9%; Score 146; DB 11; Length 653;
Best Local Similarity 47.1%; Pred. No. 6.5e-09;
Matches 32; Conservative 7; Mismatches 19; Indels 10; Gaps 3;

QY 1 CFVNGHLYREDQTSAPGLRCNLW-----DAQSLASAPVSAGNHSCYCRNPDED 52
Db 283 CFLGNGTEYRGVASTAASGLSLAWNBDLLYQELFVDS-VAAAVLLGLGLPHAYCRNPDKD 341
QY 53 PRGPWCYV 60
Db 342 ER-PWCYV 348

RESULT 13
Q8MKB1 PRELIMINARY; PRT; 564 AA.
AC Q8MKB1;
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Tissue-type plasminogen activator.
OS Oryctolagus cuniculus (Rabbit).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
OC NCBI_TaxId=9986;
RN [1]
RP SEQUENCE FROM N.A.
RA Sugiki M., Yoshida E., Anai K., Maruyama M.;
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1.
CC -1- SIMILARITY: CONTAINS 2 KRINGLE DOMAINS.
DR EMBL; AY029518; AAK40240.1; -.
DR GO; GO:0005576; C:extracellular; IEA.
DR GO; GO:0004263; F:chymotrypsin activity; IEA.
DR GO; GO:0008233; F:peptidase activity; IEA.
DR GO; GO:0004295; F:trypsin activity; IEA.
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR InterPro; IPR009003; Cys Ser trypsin.
DR InterPro; IPR006209; EGF_like.
DR InterPro; IPR000083; Fibrinctnl.

```

```
DR InterPro; IPR006210; IEFG.
DR InterPro; IPR000001; Kringle.
DR InterPro; IPR001254; Peptidase_S1.
DR InterPro; IPR001314; Peptidase_S1A.
DR Pfam; PF00008; EGF; 1.
DR Pfam; PF00039; EGF; 1.
DR Pfam; PF00051; kringle; 2.
DR Pfam; PF00089; trypsin; 1.
DR PRINTS; PR00722; CHYMOTRYPSIN.
DR PRINTS; PR00018; KRINGLE.
DR ProDom; PD000395; Kringle; 2.
DR SMART; SM00181; EGF; 1.
DR SMART; SM00058; FN1; 1.
DR SMART; SM00130; KR; 2.
DR SMART; SM00020; Tryp_SPC; 1.
DR PROSITE; PS00022; EGF_1; 1.
DR PROSITE; PS01186; EGF_2; 1.
DR PROSITE; PS01253; FIBRONECTIN_1; 1.
DR PROSITE; PS00021; KRINGLE_1; 2.
DR PROSITE; PS00070; KRINGLE_2; 2.
DR PROSITE; PS00240; TRYPSIN_DOM; 1.
DR PROSITE; PS00134; TRYPSIN_HIS; 1.
DR PROSITE; PS00135; TRYPSIN_SER; 1.
DR EGF-like domain; Glycoprotein; Hydrolase; Kringle; Protease;
KW Serine protease.
KW SEQUENCE 564 AA; 62726 MW; 459DBBAC6D4A937C CRC64;

Query Match 40.1%; Score 143; DB 6; Length 564;
Best Local Similarity 43.5%; Pred. No. 1.3e-08;
Matches 30; Conservative 7; Mismatches 20; Indels 12; Gaps 3;

Qy 1 CFWDNGHLYREDQTSAPAGRLCLNWLDAQSGLASAPVS-----GAGNHYCRNPDE 51
128 CYEDRGIGYGTWTTESGAQCWNV--NSSWLALKPYSGRKPALRLGLGNHNYCRNPDR 185
Qy 52 DPRGFWCVY 60
Db 186 DTK-PWCYV 193

RESULT 14
ID Q96SE8 PRELIMINARY; PRT; 154 AA.
AC Q96SE8;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Urokinase-type plasminogen activator amino-terminal fragment.
GN ATF.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Fu J., Bai X., Ruan C.;
RT "Cloning and expression of the amino-terminal fragment of human
RT urokinase-type plasminogen activator.";
RL Submitted (Apr-2001) to the EMBL/GenBank/DBSJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA Bai X., Fu J., Wang W., Xi X., Ruan C.;
RT "Overexpression of the amino-terminal fragment of human urokinase-type
RT plasminogen activator in breast cancer cells results in decreased
RT tumor invasion, growth and angiogenesis.";
RL Submitted (Apr-2001) to the EMBL/GenBank/DBSJ databases.
CC -!- SIMILARITY: CONTAINS 1 KRINGLE DOMAIN.
DR EMBL; AY029537; AAX38734.1; -.
DR GO; GO:0016301; F:kinase activity; IEA.
DR InterPro; IPR006209; EGF-like.
DR InterPro; IPR000001; Kringle.
DR Pfam; PF00051; kringle; 1.

Query Match 36.7%; Score 131; DB 4; Length 154;
Best Local Similarity 40.3%; Pred. No. 8.6e-08;
Matches 27; Conservative 7; Mismatches 25; Indels 8; Gaps 2;

Qy 1 CFWDNGHLYREDQTSAPAGRLCLNWLDA-----QSGLASAPVSGAGNHYCRNPDEDP 53
70 CYEGNGHYRGKASTDTMGRPCLPWNSATVLQQTYYHAHRSDALQLGLGKHYCRNPDP-NR 128
Qy 54 RGPWCYV 60
Db 129 RRPWCYV 135

RESULT 15
ID Q7ZTN9 PRELIMINARY; PRT; 709 AA.
AC Q7ZTN9;
DT 01-JUN-2003 (TrEMBLrel. 24, Created)
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Similar to macrophage stimulating 1 (Hepatocyte growth
DE factor-like).
OS Xenopus laevis (African clawed frog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipiloidea; Pipidae;
OC Xenopodinae; Xenopus.
OX NCBI_TaxID=83355;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Embryo;
RA Klein S., Strausberg R.;
RL Submitted (Jan-2003) to the EMBL/GenBank/DBSJ databases.
DR EMBL; BC044008; AAH44008.1; -.
DR GO; GO:0005509; F:calcium ion binding; IEA.
DR GO; GO:0004263; F:chymotrypsin activity; IEA.
DR GO; GO:0003809; F:thrombin activity; IEA.
DR GO; GO:0004285; F:trypsin activity; IEA.
DR GO; GO:0007596; F:blood coagulation; IEA.
DR GO; GO:0006508; F:proteolysis and peptidolysis; IEA.
DR InterPro; IPR009003; Cys_Ser_trypsin.
DR InterPro; IPR000001; Kringle.
DR InterPro; IPR003014; PAN.
DR InterPro; IPR003609; Pan_app.
DR InterPro; IPR001254; Peptidase_S1.
DR InterPro; IPR001314; Peptidase_S1A.
DR Pfam; PF00051; kringle; 4.
DR Pfam; PF00024; PAN; 1.
DR Pfam; PF00089; trypsin; 1.
DR PRINTS; PR00722; CHYMOTRYPSIN.
DR PRINTS; PR00018; KRINGLE.
DR PRINTS; PR01505; PROTHROMBIN.
DR ProDom; PD000395; Kringle; 4.
DR SMART; SM00130; KR; 4.
DR SMART; SM00473; PAN_AP; 1.
DR SMART; SM00020; Tryp_SPC; 1.
DR PROSITE; PS00021; KRINGLE_1; 4.
DR PROSITE; PS00070; KRINGLE_2; 4.
DR PROSITE; PS00240; TRYPSIN_DOM; 1.
DR SEQUENCE 709 AA; 81123 MW; 51C44D8EBD03B76A CRC64;

Query Match 36.6%; Score 130.5; DB 13; Length 709;
Best Local Similarity 38.7%; Pred. No. 5.1e-07;
```

Matches 24; Conservative 7; Mismatches 28; Indels 3; Gaps 1;  
QY 1 CFWDNGHLYREDQTSAPGLRCLNWLDAQSL---ASAPVSGAGNHSCRNPDDEPRGPW 57  
Db 370 CYHGNGELYRGRVSKTRKGMCRWEKKQNDLELSLAQPYLVPLEENTCRNPDSDSHGPW 429  
QY 58 CY 59  
Db 430 CY 431

Search completed: March 17, 2004, 07:06:56  
Job time : 11.9825 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 17, 2004, 06:55:13 ; Search time 15.3432 Seconds  
(without alignments)  
1104.909 Million cell updates/sec

Title: US-10-057-951-2\_COPY\_25\_84

Perfect score: 357

Sequence: 1 CFWDNGHLYREDQTSFAPGL.....GNHSYCRNPEDPRGPWCYV 60

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A\_Geneseq\_29Jan04:\*

1: Geneseq1980s:\*

2: Geneseq1990s:\*

3: Geneseq2000s:\*

4: Geneseq2001s:\*

5: Geneseq2002s:\*

6: Geneseq2003as:\*

7: Geneseq2003bs:\*

8: Geneseq2004s:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	357	100.0	146	5	ABR40487 Human sec
2	357	100.0	146	5	ABR40561 Human sec
3	357	100.0	263	2	Aaw87769 Human tis
4	357	100.0	263	2	Aay05219 Kringlei
5	357	100.0	263	4	Aam93748 Human pol
6	357	100.0	263	4	Aae00300 Human tis
7	357	100.0	263	5	Aau86149 Human PRO
8	357	100.0	263	5	ABR40414 Human sec
9	351	98.3	263	3	ABR43237 Human ORF
10	343	96.1	286	7	Aay05220 Kringlei
11	283	79.3	81	7	ABR42624 Human kri
12	185	52.1	56	2	Aay12615 Human 5'
13	185	51.8	55	2	Aay12397 Human 5'
14	183	51.3	39	2	Aaw72640 Nervous g
15	183	51.3	39	2	Aaw72641 Nervous g
16	149	41.7	527	2	Aaw54157 t-PA muta
17	149	41.7	527	2	Aaw54154 t-PA muta
18	148	41.5	390	1	AAP82587 Modified
19	148	41.5	483	2	AAR70879 Human tis
20	148	41.5	483	2	AAR70883 Human tis
21	148	41.5	483	2	AAR70884 Human tis
22	148	41.5	483	2	AAR70886 Human tis
23	148	41.5	483	2	AAR70878 Human tis
24	148	41.5	483	2	AAR70877 Human tis
25	148	41.5	483	2	AAR70887 Human tis

26	148	41.5	483	2	AAR70881 Human tis
27	148	41.5	483	2	AAR70882 Human tis
28	148	41.5	483	2	AAR70885 Human tis
29	148	41.5	483	2	AAR70889 Human tis
30	148	41.5	483	2	AAR70888 Human tis
31	148	41.5	483	2	AAR70890 Human tis
32	148	41.5	483	2	AAR70880 Human tis
33	148	41.5	527	2	AAR21598 tPA varia
34	148	41.5	527	2	AAR44812 Human tPA
35	148	41.5	527	2	AAR44817 Human tPA
36	148	41.5	527	2	AAR44817 Human tPA
37	148	41.5	527	2	AAR70875 Human t-P
38	148	41.5	527	2	AAR70876 Human t-P
39	148	41.5	527	2	AAR70907 Human t-P
40	148	41.5	527	2	AAR70908 Human t-P
41	148	41.5	527	2	AAR70903 Human t-P
42	148	41.5	527	2	AAR70894 Human t-P
43	148	41.5	527	2	AAR70896 Human t-P
44	148	41.5	527	2	AAR70898 Human t-P
45	148	41.5	527	2	AAR70901 Human t-P

ALIGNMENTS

RESULT 1

ABR40487  
ID ABR40487 standard; protein; 146 AA.

XX ABR40487;

XX AC

XX DT 13-JUN-2003 (first entry)

XX DE Human secreted protein #SEQ ID 237.

XX KW

XX KW Human; secreted protein; anti-HIV; neurotropic; neuroprotective;

XX KW antitaxial; immunosuppressive; immunomodulator; cytostatic; cardiac;

XX KW hepatotropic; antiinflammatory; antiallergic; antidiabetic;

XX KW gastrointestinal; antifertility; nephrotropic; virucide; hypotensive;

XX KW vasotropic; dermatological; osteopathic; antiarthritic; antiparkinsonian;

XX KW antiasthmatic; antipsoriatic; cerebroprotective; antibacterial;

XX KW fungicide; antiparasitic; protein therapy; gene therapy; cancer;

XX KW hyperproliferative disorder; leukaemia; autoimmune disorder;

XX KW immunodeficiency; HIV; infertility; uterine disorders; endometriosis;

XX KW amenorrhea; ocular disorder; neurological disorder; wound healing;

XX KW Huntington's disease; gastrointestinal disorder; inflammatory disease;

XX KW Crohn's disease; kidney disorder; renal failure; diabetes; obesity;

XX KW cirrhosis; hepatitis; cardiovascular disorder; angina; osteoarthritis;

XX KW rheumatoid arthritis; psoriasis; multiple sclerosis; asthma;

XX KW respiratory disorder; infectious disease; chromosome identification;

XX KW food additive; nutrition.

XX KW Homo sapiens.

XX WO200268628-A1.

XX PN

XX PD 06-SEP-2002.

XX XX 21-FEB-2002; 2002WO-US0053301.

XX PR 23-FEB-2001; 2001US-0270625P.

XX PR 12-JUL-2001; 2001US-0304417P.

XX PA (HUMA-) HUMAN GENOME SCI INC.

XX PI Rosen CA, Komatsoulis G, Baker KP, Fiscella M, Moore PA, Wei P;

XX PI Duan DR, Shi Y, Gupta R;

XX DR WPI; 2002-750417/81.

XX DR N-PSDB; AB232542.

XX XX New human secreted proteins and nucleic acids, useful for preventing,

XX PT treating, or ameliorating a medical condition, e.g. cancer, autoimmune

PR	23-FEB-2001; 2001US-0270625P.
PR	12-JUL-2001; 2001US-0304417P.
XX	(HUMA-) HUMAN GENOME SCI INC.
PA	Rosen CA, Komatsoulis G, Baker KP, Fiscella M, Moore PA, Wei P;
PI	Duan DR, Shi Y, Gupta R;
PP	WPI; 2002-750417/81.
DR	
XX	New human secreted proteins and nucleic acids, useful for preventing,
PT	treating, or ameliorating a medical condition, e.g. cancer, autoimmune
PT	disorder, HIV, Parkinson's disease, Alzheimer's disease, diabetes,
PT	obesity or cirrhosis.
XX	
PS	Claim 11; Page 850; 873pp; English.
XX	
CC	The invention relates to novel human secreted proteins and the genes
CC	encoding them. Genes and proteins of the invention may be useful for
CC	preventing, treating or ameliorating medical conditions e.g. by protein
CC	or gene therapy. These conditions include cancer and hyperproliferative
CC	disorders, immune cell proliferative disorders (e.g. leukaemia),
CC	autoimmune disorders, immunodeficiencies (e.g. HIV infection),
CC	infertility, placental and uterine disorders (e.g. endometriosis),
CC	amenorrhea, ocular disorders, neurological disorders (e.g. Huntington's
CC	disease), wound healing, gastrointestinal system disorders, particularly
CC	inflammatory diseases (e.g. Crohn's disease), kidney disorders (e.g.
CC	renal failure), diabetes, obesity, cirrhosis, hepatitis, cardiovascular
CC	disorders (e.g. anginal, rheumatoid arthritis, osteoarthritis, psoriasis,
CC	multiple sclerosis, respiratory disorders (e.g. asthma), or infectious
CC	diseases caused by bacterial, parasitic, viral or fungal agents. The
CC	nucleic acids are also useful for chromosome identification, radiation
CC	hybrid mapping or long-range restriction mapping. The polypeptide,
CC	poly-nucleotide, agonist or antagonist may also be used as a food additive
CC	or preservative to increase or decrease storage capabilities, fat content
CC	or other nutritional components. The sequences given in records ABR40409-
CC	ABR40590 and ABZ82464-ABZ82611 represent human secreted proteins and the
XX	genes encoding them
SQ	Sequence 146 AA;
	Query Match 100.0%; Score 357; DB 5; Length 146;
	Best Local Similarity 100.0%; Pred. No. 2.1e-34;
	Matches 50; Conservative 0; Mismatches 0; Indels 0; Gaps 0
QY	1 CFWDNCHLYREDQTSPAGRLCLNWLDAQSLGAPVSGAGNHSCYNPDDEPGPKCYV 60
Db	25 CFWDNCHLYREDQTSPAGRLCLNWLDAQSLGAPVSGAGNHSCYNPDDEPGPKCYV 84
RESULT 3	
ID	AAM87769 standard; protein; 263 AA.
XX	
AC	AAM87769;
DT	
XX	29-MAR-1999 (first entry)
DE	Human tissue plasminogen activator-like protease t-PALP.
XX	
KW	Tissue plasminogen activator-like protease; t-PALP; human;
KW	circulatory system-related disorder; blood clotting; stroke; thrombosis;
KW	peripheral arterial occlusion; pulmonary embolism; myocardiothrombosis;
KW	diagnosis; therapy.
XX	
OS	Homo sapiens.
XX	
FH	Key Location/Qualifiers
FT	Peptide 1..21
FT	/label= Sig_peptide
FT	Protein 22..263
FT	/label= Mat_protein
FT	Peptide 22..31

FT Domain /note= "epitope-bearing region"  
 FT 25..84  
 FT /note= "kringle domain"  
 FT 35..44  
 FT /note= "epitope-bearing region"  
 FT 71..81  
 FT /note= "epitope-bearing region"  
 FT 85..263  
 FT /note= "protease domain"  
 FT 91..107  
 FT /note= "epitope-bearing region"  
 FT 119..128  
 FT /note= "epitope-bearing region"  
 FT 138..147  
 FT /note= "epitope-bearing region"  
 FT 155..167  
 FT /note= "epitope-bearing region"  
 FT 193..203  
 FT /note= "epitope-bearing region"  
 FT 206..215  
 FT /note= "epitope-bearing region"  
 FT 227..237  
 FT /note= "epitope-bearing region"  
 FT 243..252  
 FT /note= "epitope-bearing region"

WO9854199-A1.

03-DEC-1998.

27-MAY-1998; 98WO-US010728.

28-MAY-1997; 97US-0048000P.

(HUMA-) HUMAN GENOME SCI INC.

Ebner R, Moore PA, Ruben SM;

WPI; 1999-070207/06.

N-PSDB; AAV99636.

New tissue plasminogen activator-like protease - useful in the diagnosis  
 FT and treatment of circulatory system-related disorders.

Claim 1; Page 56-57; 76pp; English.

XX This is the amino acid sequence of tissue plasminogen activator-like  
 CC protease (t-PALP), a novel member of the serine protease family that  
 CC shares sequence homology to human tissue plasminogen activator (see  
 CC AAW87770). The t-PALP sequence was deduced from a cDNA clone (see  
 CC AAV99636) derived from activated monocytes. The 2.5 kb t-PALP message has  
 CC also been detected in heart, brain, lung, placenta, liver, skeletal  
 CC muscle, kidney, pancreas, spleen, thymus, testis, ovary, small  
 CC intestine, colon and peripheral blood leukocytes. Isolated nucleic acids  
 CC encoding amino acids -21 to 242, -20 to 242, 1-242, 4-63 (kringle domain)  
 CC and 64-242 (protease domain) of t-PALP, or encoding epitope-bearing  
 CC portions of t-PALP, are also claimed, as are recombinant vectors, host  
 CC cells, and methods for producing t-PALP polypeptides. t-PALP may be used  
 CC to detect and treat disorders related to the circulatory system, and to  
 CC identify agonists and antagonists of t-PALP activity. The homology  
 CC between t-PALP and tPA indicates that t-PALP may be involved in the  
 CC regulation of normal and abnormal clotting in e.g. stroke, deep-vein  
 CC thrombosis, peripheral arterial occlusion, pulmonary embolism and  
 CC myocardial infarction.

XX Sequence 263 AA;

Query Match 100.0%; Score 357; DB 2; Length 263;

Best Local Similarity 100.0%; Pred. No. 3.9e-34;

Matches 60; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CFWDNGHLYREDQTSAPAGRLCLNWLDAOSGLASAPVSGAGNHSYCRNPDPRGPWCYV 60

Db 25 CFWDNGHLYREDQTSAPAGRLCLNWLDAOSGLASAPVSGAGNHSYCRNPDPRGPWCYV 84

RESULT 4

RAY05219

ID RAY05219 standard; protein; 263 AA.

XX AC RAY05219;

XX DT 17-JUN-1999 (first entry)

XX DE Kringle1 protein sequence.

XX KW Kringle1; cancer; inflammation; autoimmunity; rheumatoid arthritis; AIDS;  
 KW CNS inflammation; cerebellar degeneration; Alzheimer's disease; asthma;  
 KW Parkinson's disease; multiple sclerosis; amyotrophic lateral sclerosis;  
 KW neurological abnormality; ischaemia reperfusion injury; ischaemic injury;  
 KW cardiovascular disease; kidney disease; liver disease; aplastic anaemia;  
 KW myocardial infarction; hypotension; hypertension; allergy; infection;  
 KW myelodysplastic syndrome; haematologic abnormality; diagnosis; therapy;  
 KW male pattern baldness.

XX OS Homo sapiens.

XX PN WO9911788-A1.

XX PD 11-MAR-1999.

XX PF 02-SEP-1998; 98WO-US018270.

XX PR 02-SEP-1997; 97US-0056032P.

XX PR 01-SEP-1998; 98US-00144889.

XX PA (SMIK ) SMITHKLINE BEECHAM CORP.

XX PI Albione EP, Kikly KK;

XX DR WPI; 1999-214707/18.

XX DR N-PSDB; AAX28354.

XX PT New kringle1 polypeptides and polynucleotides.

XX PS Claim 1; Page 31-32; 42pp; English.

XX This sequence is a Kringle1 polypeptide of the invention. The kringle1  
 CC polypeptides (I) are used to screen for agonists and antagonists.  
 CC Agonists are used to treat subjects in need of enhanced activity or  
 CC expression of (I). Antagonists are used to treat subjects having need to  
 CC inhibit the activity or expression of (I). The methods can be used to  
 CC treat conditions such as cancer, inflammation, autoimmunity, allergy,  
 CC asthma, rheumatoid arthritis, CNS inflammation, cerebellar degeneration,  
 CC Alzheimer's disease, Parkinson's disease, multiple sclerosis,  
 CC amyotrophic lateral sclerosis, head injury damage and other neurological  
 CC abnormalities, ischaemia reperfusion injury, cardiovascular disease,  
 CC kidney disease, liver disease, ischaemic injury, myocardial infarction,  
 CC hypotension, hypertension, AIDS, myelodysplastic syndromes and other  
 CC haematologic abnormalities, aplastic anaemia, male pattern baldness, and  
 CC bacterial, fungal, protozoan and viral infections. The kringle1  
 CC polypeptides may also be used to generate antibodies. Determining the  
 CC presence or absence of mutations in, and analysing for the presence or  
 CC absence of expression of, kringle1 polynucleotides can be used to  
 CC diagnose a disease or susceptibility to a disease related to expression  
 CC or activity of kringle1 proteins. The polynucleotides may also be used  
 CC for chromosome identification, and mapping

XX SQ Sequence 263 AA;

Query Match 100.0%; Score 357; DB 2; Length 263;

Best Local Similarity 100.0%; Pred. No. 3.9e-34;

Matches 60; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CFWDNGHLYREDQTSAPAGRLCLNWLDAOSGLASAPVSGAGNHSYCRNPDPRGPWCYV 60



Db 25 CFWDNGHLYREDQTSAPGLRCLNWLDAQSLASAPVSGAGNHSYCRNPDEDPGRWCYV 84

RESULT 5  
AA093748  
ID AAM03748 standard; protein; 263 AA.  
XX  
AC AAM03748;  
XX  
DT 06-NOV-2001 (first entry)  
XX  
DE Human polypeptide, SEQ ID NO: 3727.  
XX  
KW Human; full length cDNA; cDNA synthesis; oligo-capping.  
XX  
OS Homo sapiens.  
XX  
PN EP1130094-A2.  
XX  
PD 05-SEP-2001.  
XX  
PF 07-JUL-2000; 2000EP-00114089.  
XX  
PR 08-JUL-1999; 99JP-00194486.  
PR 11-JAN-2000; 2000JP-00118774.  
PR 02-MAY-2000; 2000JP-00183765.  
XX  
PA (HELI-) HELIX RES INST.  
XX  
PI Ota T, Nishikawa T, Isogai T, Hayashi K, Ishii S, Kawai Y;  
PI Wakamatsu A, Sugiyama T, Nagai K, Kojima S, Otsuki T, Koga H;  
XX  
DR WPI; 2001-524255/58.  
DR N-PSDB; AAK94700.  
XX  
PT 830 Primers useful for synthesizing full length cDNA clones and their use  
PT in genetic manipulation.  
XX  
PS Claim 8; SEQ ID NO 3727; 1380pp + Sequence Listing; English.  
XX  
CC The invention relates to primers for synthesising full length cDNA  
CC clones. 830 cDNA molecules encoding a human protein have been isolated  
CC and nucleotide sequences of 5' and 3'-ends of the cDNA molecules have  
CC been determined. Primers for synthesising the full length cDNA are useful  
CC for clarifying the function of the protein encoded by the cDNA. The full  
CC length clones were obtained by construction of full length enriched cDNA  
CC libraries that were synthesised by the oligo-capping method. The primers  
CC enable the production of the full length cDNA easily without any special  
CC methods. The present sequence is a polypeptide encoded by a full length  
CC human cDNA of the invention. Note: The sequence data for this patent did  
CC not form part of the printed specification, but was obtained in CD-ROM  
CC format directly from BPO  
XX  
SQ Sequence 263 AA;  
Query Match 100.0%; Score 357; DB 4; Length 263;  
Best Local Similarity 100.0%; Pred. No. 3.9e-34;  
Matches 60; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 1 CFWDNGHLYREDQTSAPGLRCLNWLDAQSLASAPVSGAGNHSYCRNPDEDPGRWCYV 60  
Db 25 CFWDNGHLYREDQTSAPGLRCLNWLDAQSLASAPVSGAGNHSYCRNPDEDPGRWCYV 84

RESULT 6  
AAE00300  
ID AAE00300 standard; protein; 263 AA.  
XX  
AC AAE00300;  
XX  
DT 13-JUN-2001 (first entry)  
XX  
DE Human tissue-plasminogen activator-like protease (t-PALP).

XX Human; tissue-plasminogen activator-like protease; t-PALP; therapy;  
KW vascular disease; stroke; deep vein thrombosis; keloid; asthma;  
KW arterial occlusion; blood coagulation disorder; cerebroprotective;  
KW autoimmune system disorder; human immunodeficiency syndrome; cystostatic;  
KW rheumatoid arthritis; graft-versus-host disease; thyroiditis; cardiant;  
KW insulin dependent diabetes; inflammatory eye disease; allergy; antiviral;  
KW cardiovascular disease; heart disease; arrhythmia; myocardial ischaemia;  
KW hyperproliferative disorder; hypertrophic scar; neurological disease;  
KW Creutzfeldt-Jakob syndrome; neurodegenerative disorder; antiinflammatory;  
KW Alzheimer's disease; Parkinson's disease; immunosuppressive; antifungal;  
KW infectious disease; drug screening; gene therapy; neuroprotective;  
KW cancer; ophthalmological; antibacterial; vulnary.  
XX  
OS Homo sapiens.  
XX  
PH Key  
FT Binding-site  
FT 1. .165  
FT /note= "Binds to FLAG polypeptide to form t-PALP-FLAG  
FT fusion protein"  
FT Peptide  
FT 1. .21  
FT /label= Signal\_peptide  
FT Domain  
FT 4. .63  
FT /label= Kringle\_domain  
FT Region  
FT 12. .21  
FT /note= "Conserved region"  
FT Protein  
FT 22. .263  
FT /note= "Human mature tissue-plasminogen activator-like  
FT protease (t-PALP); Binds to FLAG polypeptide to form t-  
FT PALP-FLAG fusion protein"  
FT Region  
FT 22. .38  
FT /note= "Conserved region"  
FT Region  
FT 22. .31  
FT /note= "Epitope-bearing portion"  
FT Region  
FT 35. .44  
FT /note= "Epitope-bearing portion"  
FT Region  
FT 39. .49  
FT /note= "Conserved region"  
FT Region  
FT 50. .62  
FT /note= "Conserved region"  
FT Region  
FT 63. .84  
FT /note= "Conserved region"  
FT Domain  
FT 64. .242  
FT /label= Protease\_domain  
FT Region  
FT 71. .81  
FT /note= "Epitope-bearing portion"  
FT Region  
FT 85. .97  
FT /note= "Conserved region"  
FT Region  
FT 91. .107  
FT /note= "Epitope-bearing portion"  
FT Region  
FT 100. .118  
FT /note= "Conserved region"  
FT Region  
FT 119. .128  
FT /note= "Epitope-bearing portion"  
FT Region  
FT 119. .127  
FT /note= "Conserved region"  
FT Region  
FT 128. .143  
FT /note= "Conserved region"  
FT Region  
FT 138. .147  
FT /note= "Epitope-bearing portion"  
FT Region  
FT 146. .163  
FT /note= "Conserved region"  
FT Region  
FT 155. .167  
FT /note= "Epitope-bearing portion"  
FT Region  
FT 164. .180  
FT /note= "Conserved region"  
FT Region  
FT 186. .200  
FT /note= "Conserved region"  
FT Region  
FT 193. .203  
FT /note= "Epitope-bearing portion"  
FT Region  
FT 201. .220  
FT /note= "Conserved region"  
FT Region  
FT 206. .215

FT /note= "Epitope-bearing portion"  
 FT 221..236  
 FT /note= "Conserved region"  
 FT 227..237  
 FT /note= "Epitope-bearing portion"  
 FT 237..248  
 FT /note= "Conserved region"  
 FT 243..252  
 FT /note= "Epitope-bearing portion"  
 FT 249..263  
 FT /note= "Conserved region"  
 FT  
 FT  
 FT WO200125252-A1.  
 FT  
 FT 12-APR-2001.  
 FT  
 FT 03-OCT-2000; 2000WO-US027239.  
 FT  
 FT 04-OCT-1999; 99US-00411977.  
 FT  
 FT (HUMA-) HUMAN GENOME SCI INC.  
 FT  
 FT Moore PA, Ruben SM, Ebner R;  
 FT  
 FT WPI; 2001-235402/24.  
 FT N-PSDB; AAD03460.  
 FT  
 FT New (gene encoding and antibody immunospecific for a) tissue-plasminogen  
 FT activator-like protease, useful for the diagnosis and treatment of  
 FT (cardio)vascular diseases, hyperproliferative disorders, immune system  
 FT disorders and cancers.  
 FT  
 FT Claim 17; Fig 1; 323pp; English.  
 FT  
 FT The present amino acid sequence is HMSIB42 clone human tissue-plasminogen  
 FT activator-like protease (t-PALP). The t-PALP sequence and their  
 FT (ant)agonists are useful for the diagnosis and treatment of vascular  
 FT diseases e.g. stroke, deep vein thrombosis and arterial occlusion, blood  
 FT coagulation disorders, (auto)immune system disorders e.g. human  
 FT immunodeficiency syndrome, rheumatoid arthritis, graft-versus-host  
 FT disease, thyroiditis, insulin dependent diabetes and inflammatory eye  
 FT disease, allergic reactions e.g. asthma, cardiovascular diseases e.g.  
 FT heart disease, arrhythmia and myocardial ischaemia, hyperproliferative  
 FT disorders, cancers, hypertrophic scars and keloids, neurological diseases  
 FT e.g. Creutzfeldt-Jakob syndrome, neurodegenerative disorders e.g.  
 FT Alzheimer's disease and Parkinson's disease and infectious disease e.g.  
 FT viral, bacterial and fungal infections. The t-PALP sequences are also  
 FT useful for drug screening. The t-PALP nucleotides are useful as  
 FT chromosome markers and are involved in gene therapy  
 FT  
 FT SQ Sequence 263 AA;  
 FT  
 FT Query Match 100.0%; Score 357; DB 4; Length 263;  
 FT Best Local Similarity 100.0%; Pred. No. 3.9e-34;  
 FT Matches 60; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 FT  
 FT QY 1 CFWDNGHLYREDQTSAPAGRLCLNWLDAQSLASAPVSGAGNHSYCRNPDEDPGWCYV 60  
 FT DB 25 CFWDNGHLYREDQTSAPAGRLCLNWLDAQSLASAPVSGAGNHSYCRNPDEDPGWCYV 84  
 FT  
 FT RESULT 7  
 FT AAU86149  
 FT  
 FT AC AAU86149;  
 FT  
 FT 15-JUL-2002 (first entry)  
 FT  
 FT Human PRO264 polypeptide.  
 FT  
 FT Human, PRO; benign tumour, malignant tumour; lymphoid malignancy;  
 FT leukaemia; neuronal disorder; stromal disorder; blastocoelec disorder;  
 FT

KW inflammatory disorder; immune disorder; angiogenic disorder; cytostatic;  
 KW neuroprotective.  
 OS Homo sapiens.  
 PN WO200153486-A1.  
 XX 26-JUL-2001.  
 XX  
 XX 11-FEB-2000; 2000WO-US003565.  
 XX  
 XX 08-MAR-1999; 99WO-US005028.  
 XX 11-MAR-1999; 99US-0123972P.  
 XX 11-MAY-1999; 99US-0133459P.  
 XX 02-JUN-1999; 99WO-US012252.  
 XX 22-JUN-1999; 99US-0140650P.  
 XX 22-JUN-1999; 99US-0140653P.  
 XX 26-JUL-1999; 99US-0144758P.  
 XX 26-JUL-1999; 99US-0145698P.  
 XX 28-JUL-1999; 99US-0146222P.  
 XX 17-AUG-1999; 99US-0149395P.  
 XX 31-AUG-1999; 99US-0151689P.  
 XX 01-SEP-1999; 99WO-US020111.  
 XX 15-SEP-1999; 99WO-US021090.  
 XX 30-NOV-1999; 99WO-US028313.  
 XX 01-DEC-1999; 99WO-US028301.  
 XX 01-DEC-1999; 99WO-US028634.  
 XX 05-JAN-2000; 2000WO-US000219.  
 XX  
 XX (GETH ) GENENTECH INC.  
 XX  
 XX Ashkenazi AJ, Goddard A, Godowski PJ, Gurney AL, Hillan KJ;  
 XX Marsters SA, Pan J, Pitti RM, Roy MA, Smith V, Stone DM;  
 XX Watanabe CK, Wood WI;  
 XX  
 XX WPI; 2002-205567/26.  
 XX N-PSDB; ABK40275.  
 XX  
 XX Thirty five nucleic acids encoding PRO polypeptides, useful for treating  
 XX benign or malignant tumors, leukemias and lymphoid malignancies,  
 XX inflammatory, angiogenic and immunologic disorders.  
 XX  
 XX Claim 61; Fig 44; 302pp; English.  
 XX  
 XX The present invention relates to the isolation of novel human PRO  
 XX polypeptides and the polynucleotide sequences encoding them. The PRO  
 XX polypeptides, agonists, antagonists or anti-PRO antibodies are useful for  
 XX treating benign or malignant tumors (e.g. renal, kidney, bladder,  
 XX breast, etc), leukemias and lymphoid malignancies, other disorders such  
 XX as neuronal, glial, astrocytal, hypothalamic, glandular, macrophagal,  
 XX stromal and blastocoelec disorders, inflammatory, immune and angiogenic  
 XX disorders. The polynucleotide sequences are also useful in gene therapy.  
 XX AAU86128-AAU86162 represent the human PRO polypeptides of the invention  
 XX  
 XX SQ Sequence 263 AA;  
 XX  
 XX Query Match 100.0%; Score 357; DB 5; Length 263;  
 XX Best Local Similarity 100.0%; Pred. No. 3.9e-34;  
 XX Matches 60; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 XX  
 XX QY 1 CFWDNGHLYREDQTSAPAGRLCLNWLDAQSLASAPVSGAGNHSYCRNPDEDPGWCYV 60  
 XX DB 25 CFWDNGHLYREDQTSAPAGRLCLNWLDAQSLASAPVSGAGNHSYCRNPDEDPGWCYV 84  
 XX  
 XX RESULT 8  
 XX ABR40414  
 XX ID ABR40414 standard; protein; 263 AA.  
 XX  
 XX AC ABR40414;  
 XX  
 XX 13-JUN-2003 (first entry)  
 XX  
 XX

DE Human secreted protein #SEQ ID 164.

XX Human; secreted protein; anti-HIV; nontropic; neuroprotective;  
 KW antitumoral; immunosuppressive; immunomodulator; cytostatic; cardiant;  
 KW hepatotropic; antinflammatory; antiallergic; antidiabetic;  
 KW gastrointestinal; antifertility; nephrotropic; virucide; hypotensive;  
 KW vasotropic; dermatological; osteopathic; antiarthritic; antiparkinsonian;  
 KW antitastmatic; antipariatic; cerebroprotective; antibacterial;  
 KW fungicide; antiparasitic; protein therapy; gene therapy; cancer;  
 KW hyperproliferative disorder; leukaemia; autoimmune disorder;  
 KW immunodeficiency; HIV; infertility; uterine disorders; endometriosis;  
 KW amenorrhea; ocular disorder; neurological disorder; wound healing;  
 KW Huntington's disease; gastrointestinal disorder; inflammatory disease;  
 KW Crohn's disease; kidney disorder; renal failure; diabetes; obesity;  
 KW cirrhosis; hepatitis; cardiovascular disorder; angina; osteoarthritis;  
 KW rheumatoid arthritis; psoriasis; multiple sclerosis; asthma;  
 KW respiratory disorder; infectious disease; chromosome identification;  
 KW food additive; nutrition.

XX Homo sapiens.

OS WO200268628-A1.

PN 06-SEP-2002.

PD 21-FEB-2002; 2002WO-US005301.

PF 23-FEB-2001; 2001US-0270625P.

PR 12-JUL-2001; 2001US-0304417P.

XX (HUMA-) HUMAN GENOME SCI INC.

PA Rosen CA, Komatsoulis G, Baker KP, Fiscella M, Moore PA, Wei P;  
 PI Duan DR, Shi Y, Gupta R;  
 XX WPI: 2002-750417/81.  
 DR N-PSDB; ABZ82459.

XX New human secreted proteins and nucleic acids, useful for preventing,  
 PT treating, or ameliorating a medical condition, e.g. cancer, autoimmune  
 PT disorder, HIV, Parkinson's disease, Alzheimer's disease, diabetes,  
 PT obesity or cirrhosis.

XX Claim 11; Page 755; 873pp; English.

PS The invention relates to novel human secreted proteins and the genes  
 XX encoding them. Genes and proteins of the invention may be useful for  
 CC preventing, treating or ameliorating medical conditions e.g. by protein  
 CC or gene therapy. These conditions include cancer and hyperproliferative  
 CC disorders, immune cell proliferative disorders (e.g. leukaemia),  
 CC autoimmune disorders, immunodeficiencies (e.g. HIV infection),  
 CC infertility, placental and uterine disorders (e.g. endometriosis),  
 CC amenorrhea, ocular disorders, neurological disorders (e.g. Huntington's  
 CC disease), wound healing, gastrointestinal system disorders, particularly  
 CC inflammatory diseases (e.g. Crohn's disease), kidney disorders (e.g.  
 CC renal failure), diabetes, obesity, cirrhosis, hepatitis, cardiovascular  
 CC disorders (e.g. angina), rheumatoid arthritis, osteoarthritis, psoriasis,  
 CC multiple sclerosis, respiratory disorders (e.g. asthma) or infectious  
 CC diseases caused by bacterial, parasitic, viral or fungal agents. The  
 CC nucleic acids are also useful for chromosome identification, radiation  
 CC hybrid mapping or long-range restriction mapping. The polypeptide,  
 CC or polynucleotide, agonist or antagonist may also be used as a food additive  
 CC or preservative to increase or decrease storage capabilities, fat content  
 CC or other nutritional components. The sequences given in records ABR40409-  
 CC ABR40590 and ABZ82464-ABZ82611 represent human secreted proteins and the  
 CC genes encoding them

XX Sequence 263 AA;

XX Query Match 100.0%; Score 357; DB 5; Length 263;  
 XX Best Local Similarity 100.0%; Pred. No. 3,9e-34;  
 XX Matches 60; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CFWDNGHLYREDQTSAPGLRCLNWLDAQSGLASAPVSGAGNHSYCRNPDEPRGWCYV 60  
 DB 25 CFWDNGHLYREDQTSAPGLRCLNWLDAQSGLASAPVSGAGNHSYCRNPDEPRGWCYV 84

RESULT 9

AB43237

ID AAB43237 standard; protein; 263 AA.

XX AAB43237;

XX 08-FEB-2001 (first entry)

DE Human ORFX ORF3001 polypeptide sequence SEQ ID NO:6002.

XX Human; open reading frame; ORFX; detection; cytostatic; hepatotropic;  
 KW vulnery; antipsoriatic; antiparkinsonian; nontropic; neuroprotective;  
 KW anticonvulsant; osteopathic; antiarthritic; immunosuppressant; cardiant;  
 KW immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;  
 KW hypotensive; dermatological; immunosuppressive; antinflammatory;  
 KW antiviral; antibacterial; antifungal; antirheumatic; antichyroid;  
 KW antianaemic; gene therapy; cancer; proliferative disorder; hypertension;  
 KW neurodegenerative disorder; osteoarthritis; graft vs host disease;  
 KW cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;  
 KW cholesterol ester storage; systemic lupus erythematosus; infection;  
 KW severe combined immunodeficiency; malaria; autoimmune disorder; asthma;  
 KW allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;  
 KW bone damage; cartilage damage; antiinflammatory disease; coagulation;  
 KW thrombosis; contraceptive.

XX Homo sapiens.

OS WO200058473-A2.

PN 05-OCT-2000.

PD 31-MAR-2000; 2000WO-US008621.

PF 31-MAR-1999; 99US-0127607P.

PR 02-APR-1999; 99US-0127636P.

PR 05-APR-1999; 99US-0127728P.

PR 30-MAR-2000; 2000US-00540763.

XX (CURA-) CURAGEN CORP.

PA Shimkets RA, Leach M;  
 PI WPI: 2000-602362/57.  
 DR N-PSDB; AAC77445.

XX Novel nucleic acids and peptides derived from open reading frame X,  
 PT useful for treating e.g. cancers, proliferative disorders,  
 PT neurodegenerative disorders and cardiovascular disease.

XX Claim 11; Page 5181-5182; 5507pp; English.

XX AAC77446 to AAC77606 encode the proteins given in AAB40237 to AAB43397,  
 CC which represent the human ORFX open reading frames 1 to 3161. The ORFX  
 CC sequences have activities such as: cytostatic; hepatotropic; vulnery;  
 CC antipsoriatic; antiparkinsonian; nontropic; neuroprotective; osteopathic;  
 CC anticonvulsant; antiarthritic; immunosuppressant; immunostimulant;  
 CC cardiant; thrombolytic; coagulant; vasotropic; antidiabetic; hypotensive;  
 CC dermatological; immunosuppressive; antinflammatory; antibacterial;  
 CC antiviral; antifungal; antirheumatic; antichyroid; and antianaemic. The  
 CC sequences can be used for determining the presence of or predisposition  
 CC to, or preventing or treating pathological conditions associated with an  
 CC ORFX-associated disorder. The nucleic acids can be used to express ORFX  
 CC proteins in gene therapy vectors. The proteins and nucleic acids may be  
 CC used to treat cancers, proliferative disorders, neurodegenerative  
 CC disorders, osteoarthritis, graft vs host disease, cardiovascular disease,  
 CC diabetes mellitus, hypertension, hypothyroidism, cholesterol ester  
 CC storage, systemic lupus erythematosus, severe combined immunodeficiency  
 CC (SCID), AIDS, viral, bacterial or fungal infection, malaria, autoimmune

CC disorders, asthma, allergies, aplastic anaemia, burns, wounds, bone and  
CC cartilage damage, nocturnal haemoglobinuria, antiinflammatory disease; to  
CC enhance coagulation; to inhibit thrombosis; and as a contraceptive  
XX  
XX  
SQ Sequence 263 AA;

Query Match 98.3%; Score 351; DB 3; Length 263;  
Best Local Similarity 98.3%; Pred. No. 2e-33;  
Matches 59; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CFWDNGHLYREDTSPAPGLRCLNWLDAQSLASAPVSGAGNHSYCRNPDEDPGPGWCYV 60  
|||  
DB 25 CFWDNGHLYREDTSPAPGLRCLNWLDAQSLASAPVSGAGNHSYCRNPDEDPGPGWCYV 84  
|||

RESULT 10  
AAY05220  
ID AAY05220 standard; protein; 286 AA.  
XX  
AC AAY05220;  
XX  
DT 17-JUN-1999 (first entry)  
XX  
DE Kringle1 protein sequence.  
XX  
KW Kringle1; cancer; inflammation; autoimmunity; rheumatoid arthritis; AIDS;  
KW CNS inflammation; cerebellar degeneration; Alzheimer's disease; asthma;  
KW Parkinson's disease; multiple sclerosis; amyotrophic lateral sclerosis;  
KW neurological abnormality; ischaemia reperfusion injury; ischaemic injury;  
KW cardiovascular disease; kidney disease; liver disease; aplastic anaemia;  
KW myocardial infarction; hypotension; hypertension; allergy; infection;  
KW myelodysplastic syndrome; haematologic abnormality; diagnosis; therapy;  
KW male pattern baldness.  
XX  
OS Homo sapiens.  
XX  
PN WO9911788-A1.  
XX  
PD 11-MAR-1999.  
XX  
PP 02-SEP-1998; 98WO-US018270.  
XX  
PR 02-SEP-1997; 97US-0056032P.  
PR 01-SEP-1998; 98US-00144889.  
XX  
PA (SMIX ) SMITHKLINE BEECHAM CORP.  
XX  
PI Albone EF, Kikly KK;  
XX  
DR WPI; 1999-214707/18.  
DR N-PSDS; AAX28355.  
XX  
FT New kringle1 polypeptides and polynucleotides.  
XX  
PS Claim 14; Page 33; 42pp; English.  
XX  
CC This sequence is a Kringle1 polypeptide of the invention. The kringle1  
CC polypeptides (I) are used to screen for agonists and antagonists.  
CC Agonists are used to treat subjects in need of enhanced activity or  
CC expression of (I). Antagonists are used to treat subjects having need to  
CC inhibit the activity or expression of (I). The methods can be used to  
CC treat conditions such as cancer, inflammation, autoimmunity, allergy,  
CC asthma, rheumatoid arthritis, CNS inflammation, cerebellar degeneration,  
CC Alzheimer's disease, Parkinson's disease, multiple sclerosis,  
CC amyotrophic lateral sclerosis, head injury damage and other neurological  
CC abnormalities, ischaemia reperfusion injury, cardiovascular disease,  
CC kidney disease, liver disease, ischaemic injury, myocardial infarction,  
CC hypotension, hypertension, AIDS, myelodysplastic syndromes and other  
CC haematologic abnormalities, aplastic anaemia, male pattern baldness, and  
CC bacterial, fungal, protozoan and viral infections. The kringle1  
CC polypeptides may also be used to generate antibodies. Determining the  
CC presence or absence of mutations in, and analysing for the presence or  
CC absence of expression of, kringle1 polynucleotides can be used to

CC diagnose a disease or susceptibility to a disease related to expression  
CC or activity of kringle1 proteins. The polynucleotides may also be used  
CC for chromosome identification, and mapping  
XX  
XX  
SQ Sequence 286 AA;

Query Match 96.1%; Score 343; DB 2; Length 286;  
Best Local Similarity 96.7%; Pred. No. 2e-32;  
Matches 58; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 CFWDNGHLYREDTSPAPGLRCLNWLDAQSLASAPVSGAGNHSYCRNPDEDPGPGWCYV 60  
|||  
DB 25 CFWDNGHLYREDTSPAPGLRCLNWLDAQSGPASAPVSGADNHSYCRNPDEDPGPGWCYV 84  
|||

RESULT 11  
ABR42624  
ID ABR42624 standard; protein; 81 AA.  
XX  
AC ABR42624;  
XX  
DT 26-AUG-2003 (first entry)  
XX  
DE Human kringle containing protein.  
XX  
KW Human; abrogen; kringle; angiogenesis; inhibitor; tumour; metastasis;  
KW cyrostatic; gene therapy; expressed sequence tag; EST.  
XX  
OS Homo sapiens.  
XX  
FH Key Location/Qualifiers  
FT Misc-difference 4 /note= "the identity of this residue is unclear in Fig 2"  
FT Misc-difference 6 /note= "the identity of this residue is unclear in Fig 2"  
FT Misc-difference 7 /note= "the identity of this residue is unclear in Fig 2"  
FT Misc-difference 16 /note= "the identity of this residue is unclear in Fig 2"  
FT Misc-difference 29 /note= "the identity of this residue is unclear in Fig 2"  
FT Misc-difference 32 /note= "the identity of this residue is unclear in Fig 2"  
FT Misc-difference 46 /note= "the identity of this residue is unclear in Fig 2"  
FT Domain 50..56 /note= "kringle domain"  
FT Misc-difference 57 /note= "the identity of this residue is unclear in Fig 2"  
FT Misc-difference 62 /note= "the identity of this residue is unclear in Fig 2"  
FT Misc-difference 70 /note= "the identity of this residue is unclear in Fig 2"  
FT Misc-difference 72 /note= "the identity of this residue is unclear in Fig 2"  
FT Misc-difference 77 /note= "the identity of this residue is unclear in Fig 2"  
FT  
XX WO2003042354-A2.  
XX  
XX 22-MAY-2003.  
XX  
XX 04-SEP-2002; 2002WO-US027885.  
XX  
XX 04-SEP-2001; 2001US-0316300P.  
XX (AVET ) AVENTIS PHARM INC.  
XX Nesbit M, Fong TC, Brockstedt D;  
XX WPI; 2003-449566/42.  
XX  
XX New abrogen polypeptide, useful for treating an angiogenesis related

PT diseases e.g. tumor metastasis.  
 XX  
 PS Disclosure; Fig 2; 95pp; English.  
 XX  
 CC The present sequence is the protein sequence of a hypothetical kringle-  
 CC containing protein encoded by an expressed sequence tag. The invention  
 CC relates to novel abrogen polypeptides that are derived from kringle-  
 CC containing proteins. The abrogens are potent inhibitors of endothelial  
 CC proliferation and angiogenesis. They are capable of inhibiting or  
 CC reducing cell proliferation induced by both basic fibroblast growth  
 CC factor and vascular endothelial growth factor in a specific endothelial  
 CC cell proliferation assay. Vectors that expressed abrogen polypeptides in  
 CC vivo were shown to reduce tumour metastasis in 2 lung cancer models. The  
 CC invention provides abrogen polypeptides and polynucleotides, and methods  
 CC of using these to treat an angiogenesis-related disease or disorder, e.g.  
 CC tumour metastasis (claimed)  
 XX  
 XX Sequence 81 AA;  
 SQ  
 Query Match 79.3%; Score 283; DB 7; Length 81;  
 Best Local Similarity 83.1%; Pred. No. 6.9e-26;  
 Matches 49; Conservative 0; Mismatches 10; Indels 0; Gaps 0;  
 QY 1 CFWDNGHLYREDQTSFAPGLRCLNWLDAQSLASAPVSGAGNHXYCRNFDPRPCWY 59  
 Db 3 CAWXXGHLYREDQTSFAPGLRCLNWLDAQSLASAPVSGAGNHXYCRNFDPRPCWY 61  
 RESULT 12  
 AAY12615  
 ID AAY12615 standard; protein; 56 AA.  
 XX  
 AC AAY12615;  
 XX  
 DT 22-JUN-1999 (first entry)  
 XX  
 DE Human 5' EST secreted protein SEQ ID NO: 280 from WO 9906553.  
 XX  
 KW Human; secreted protein; EST; expressed sequence tag; diagnosis;  
 KW forensic; gene therapy; chromosome mapping; signal peptide;  
 KW upstream regulatory sequence; cytokine activity; cell proliferation;  
 KW differentiation; haematopoiesis regulation; tissue growth regulation;  
 KW reproductive hormone regulation; chemotactic; chemokinetic; haemostatic;  
 KW thrombolytic; anti-inflammatory; tumour inhibition; antitumour.  
 XX  
 OS Homo sapiens.  
 XX  
 FN WO9906553-A2.  
 XX  
 PD 11-FEB-1999.  
 XX  
 PF 31-JUL-1998; 98WO-IB001237.  
 XX  
 PR 01-AUG-1997; 97US-00905051.  
 XX  
 PA (GBST ) GENSET.  
 XX  
 PI Dumas Milne Edwards J, Duclert A, Lacroix B;  
 XX  
 DR WPI; 1999-153783/13.  
 DR N-PSDB; AAX41473.  
 XX  
 PT New nucleic acids encoding human secreted proteins - obtained from cDNA  
 PT libraries derived from umbilical cord, lymph ganglia, lymphocytes and  
 PT placental tissue.  
 PS Claim 34; Page 376; 41pp; English.  
 XX  
 CC AAX41379 to AAX41526 represent 5' expressed sequence tags (ESTs) for  
 CC human secreted proteins, and encode the proteins given in AAY12521 to  
 CC AAY12668, respectively. The proteins given represent the signal peptide  
 CC and an N-terminal fragment of a secreted protein. The nucleic acid  
 CC sequences can be used for producing secreted human gene products. They

CC can also be used to develop products for diagnosis and therapy. The  
 CC proteins obtained may have cytokine activity, cell  
 CC proliferation/differentiation activity, haematopoiesis regulating  
 CC activity, tissue growth regulating activity, reproductive hormone  
 CC regulating activity, chemotactic/ chemokinetic activity, haemostatic and  
 CC thrombolytic activity, receptor/ ligand activity, antiinflammatory  
 CC activity, tumour inhibition activity or other activities. The products  
 CC can be used in forensic, gene therapy and chromosome mapping procedures.  
 CC The sequences can also be used for obtaining corresponding promoter.  
 CC sequences. The nucleic acids encoding the signal peptide can be used for  
 CC directing extracellular secretion of a polypeptide or the insertion of a  
 CC polypeptide into a membrane, or importing a polypeptide into a cell  
 XX  
 XX Sequence 56 AA;  
 SQ  
 Query Match 52.1%; Score 186; DB 2; Length 56;  
 Best Local Similarity 96.9%; Pred. No. 1.6e-14;  
 Matches 31; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 CFWDNGHLYREDQTSFAPGLRCLNWLDAQSL 32  
 Db 25 CFWDNGHLYREDQTSFAPGLRCLNWLDAQSL 56  
 RESULT 13  
 AAY12397  
 ID AAY12397 standard; protein; 55 AA.  
 XX  
 AC AAY12397;  
 XX  
 DT 17-JUN-1999 (first entry)  
 XX  
 DE Human 5' EST secreted protein SEQ ID NO:428.  
 XX  
 KW Human; secreted protein; EST; expressed sequence tag; diagnosis;  
 KW forensic; gene therapy; chromosome mapping; signal peptide;  
 KW upstream regulatory sequence; cytokine activity; cell proliferation;  
 KW differentiation; haematopoiesis regulation; tissue growth regulation;  
 KW reproductive hormone regulation; chemotactic; chemokinetic; haemostatic;  
 KW thrombolytic; anti-inflammatory; tumour inhibition.  
 XX  
 OS Homo sapiens.  
 XX  
 FN WO9906548-A2.  
 XX  
 PD 11-FEB-1999.  
 XX  
 PF 31-JUL-1998; 98WO-IB001222.  
 XX  
 PR 01-AUG-1997; 97US-00905135.  
 XX  
 PA (GBST ) GENSET.  
 XX  
 PI Dumas Milne Edwards J, Duclert A, Lacroix B;  
 XX  
 DR WPI; 1999-153778/13.  
 DR N-PSDB; AAX41230.  
 XX  
 PT New nucleic acids encoding human secreted proteins - obtained from cDNA  
 PT libraries prepared from e.g. liver, ovary, brain, prostate, kidney, lung,  
 PT umbilical cord, placenta and colon tissue.  
 XX  
 PS Claim 27; Page 744; 824pp; English.  
 XX  
 CC AAX41094 to AAX41347 represent 5' expressed sequence tags (ESTs) for  
 CC human secreted proteins, and encode the proteins given in AAY12261 to  
 CC AAY12514, respectively. The proteins given represent the signal peptide  
 CC and an N-terminal fragment of a secreted protein. The nucleic acid  
 CC sequences can be used for producing secreted human gene products. They  
 CC can also be used to develop products for diagnosis and therapy. The  
 CC proteins obtained may have cytokine activity, cell  
 CC proliferation/differentiation activity, haematopoiesis regulating  
 CC activity, tissue growth regulating activity, reproductive hormone

CC regulating activity, chemotactic/ chemokinetic activity, haemostatic and  
CC thrombolytic activity, receptor/ ligand activity, anti-inflammatory  
CC activity, tumour inhibition activity or other activities. The products  
CC can be used in forensic, gene therapy and chromosome mapping procedures.  
CC The sequences can also be used for obtaining corresponding promoter  
CC sequences. The nucleic acids encoding the signal peptide can be used for  
CC directing extracellular secretion of a polypeptide or the insertion of a  
CC polypeptide into a membrane, or importing a polypeptide into a cell  
XX  
SQ Sequence 55 AA;  
  
Query Match 51.8%; Score 185; DB 2; Length 55;  
Best Local Similarity 100.0%; Pred. No. 2e-14; Indels 0; Gaps 0;  
Matches 31; Conservative 0; Mismatches 0;  
  
QY 1 CFWDNGHLYREDQTSAPGLRCLNWLDAQSG 31  
DB 25 CFWDNGHLYREDQTSAPGLRCLNWLDAQSG 55  
  
RESULT 14  
AAW72640  
ID AAW72640 standard; peptide; 39 AA.  
XX  
AC AAW72640;  
XX  
DT 05-JAN-1999 (first entry)  
XX  
DE Nervous glia cell growth factor N-terminal peptide #1.  
XX  
KW Nervous glia cell growth factor; human; urine; secretion promoter;  
KW choline acetyltransferase activity enhancer; nervous disease.  
XX  
OS Homo sapiens.  
XX  
FH Key Location/Qualifiers  
FT Misc-difference 25 /note= "unspecified"  
FT Misc-difference 29 /note= "unspecified"  
FT  
FT  
XX JPI0265498-A.  
XX  
PD 06-OCT-1998.  
XX  
PF 24-MAR-1997; 97JP-00090305.  
XX  
PR 24-MAR-1997; 97JP-00090305.  
XX  
PA (NICH-) JAPAN CHEM RES CO LTD.  
XX  
DR WPI; 1998-589719/50.  
XX  
PT Nervous glia cell growth factor derived from human urine - used for  
PT treatment of nervous diseases.  
XX  
PS Claim 2; Fig 6; 14pp; Japanese.  
XX  
CC The present invention describes nervous glia cell growth factor, which is  
CC purified from human urine by ultrafiltration, salting-out by ammonium  
CC sulphate, gel filtration, ion exchange chromatography and reversed phase  
CC chromatography, and has a M.W. of 29 kDa by sodium dodecyl sulphate-  
CC polyacrylamide gel electrophoresis. Also described are: (1) a secretion  
CC promoter for the nerve growth factor of glia cell consisting of the above  
CC growth factor, an enhancer for choline acetyltransferase activity of  
CC neuron consisting of the above growth factor; and (2) DNA encoding  
CC nervous glia growth factor containing a DNA sequence coding the amino  
CC acid sequence shown by the two 39 amino acid sequences as given in  
CC AAW72640 and AAW72641 which are identical, except one starts with Tyr  
CC and the other with Ser (i.e. they are from different DNA transcripts).  
CC The glia cell growth factor can be prepared in a large amount and the  
CC factor can be used for the treatment of nervous diseases  
XX

SQ Sequence 39 AA;  
  
Query Match 51.3%; Score 183; DB 2; Length 39;  
Best Local Similarity 94.3%; Pred. No. 2.4e-14;  
Matches 33; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
  
QY 2 FWDNGHLYREDQTSAPGLRCLNWLDAQSGLASAP 36  
DB 5 FWDNGHLYREDQTSAPGLRCLNWLDAQSGLASAP 39  
  
RESULT 15  
AAW72641  
ID AAW72641 standard; peptide; 39 AA.  
XX  
AC AAW72641;  
XX  
DT 05-JAN-1999 (first entry)  
XX  
DE Nervous glia cell growth factor N-terminal peptide #2.  
XX  
KW Nervous glia cell growth factor; human; urine; secretion promoter;  
KW choline acetyltransferase activity enhancer; nervous disease.  
XX  
OS Homo sapiens.  
XX  
FH Key Location/Qualifiers  
FT Misc-difference 25 /note= "unspecified"  
FT Misc-difference 29 /note= "unspecified"  
FT  
FT  
XX JPI0265498-A.  
XX  
PD 06-OCT-1998.  
XX  
PF 24-MAR-1997; 97JP-00090305.  
XX  
PR 24-MAR-1997; 97JP-00090305.  
XX  
PA (NICH-) JAPAN CHEM RES CO LTD.  
XX  
DR WPI; 1998-589719/50.  
XX  
PT Nervous glia cell growth factor derived from human urine - used for  
PT treatment of nervous diseases.  
XX  
PS Claim 3; Fig 7; 14pp; Japanese.  
XX  
CC The present invention describes nervous glia cell growth factor, which is  
CC purified from human urine by ultrafiltration, salting-out by ammonium  
CC sulphate, gel filtration, ion exchange chromatography and reversed phase  
CC chromatography, and has a M.W. of 29 kDa by sodium dodecyl sulphate-  
CC polyacrylamide gel electrophoresis. Also described are: (1) a secretion  
CC promoter for the nerve growth factor of glia cell consisting of the above  
CC growth factor, an enhancer for choline acetyltransferase activity of  
CC neuron consisting of the above growth factor; and (2) DNA encoding  
CC nervous glia growth factor containing a DNA sequence coding the amino  
CC acid sequence shown by the two 39 amino acid sequences as given in  
CC AAW72640 and AAW72641, which are identical, except one starts with Tyr  
CC and the other with Ser (i.e. they are from different DNA transcripts).  
CC The glia cell growth factor can be prepared in a large amount and the  
CC factor can be used for the treatment of nervous diseases  
XX  
SQ Sequence 39 AA;  
  
Query Match 51.3%; Score 183; DB 2; Length 39;  
Best Local Similarity 94.3%; Pred. No. 2.4e-14;  
Matches 33; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
  
QY 2 FWDNGHLYREDQTSAPGLRCLNWLDAQSGLASAP 36  
DB 5 FWDNGHLYREDQTSAPGLRCLNWLDAQSGLASAP 39

us-10-057-951-2\_copy\_25\_84.rag

Thu Mar 18 15:28:05 2004

Search completed: March 17, 2004, 07:03:43  
Job time : 16.3432 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.  
OM protein - protein search, using sw model  
Run on: March 18, 2004, 13:12:33 ; Search time 10.9017 Seconds  
(without alignments)  
1425.213 Million cell updates/sec

Title: US-10-057-951-2\_COPY\_25\_84  
Perfect score: 357  
Sequence: 1 CFWDNGHLYREDTSPAGL.....GNHSYCRNPDEPRGWCYV 60

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1049977 seqs, 258955339 residues  
Total number of hits satisfying chosen parameters: 1049977

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Published Applications AA:

- 1: /cgn2\_6/ptodata/2/pubpaa/US07\_PUBCOMB.pep.\*
- 2: /cgn2\_6/ptodata/2/pubpaa/PCT\_NEW\_PUB.pep.\*
- 3: /cgn2\_6/ptodata/2/pubpaa/US06\_NEW\_PUB.pep.\*
- 4: /cgn2\_6/ptodata/2/pubpaa/US06\_PUBCOMB.pep.\*
- 5: /cgn2\_6/ptodata/2/pubpaa/US07\_NEW\_PUB.pep.\*
- 6: /cgn2\_6/ptodata/2/pubpaa/PCTUS\_PUBCOMB.pep.\*
- 7: /cgn2\_6/ptodata/2/pubpaa/US08\_NEW\_PUB.pep.\*
- 8: /cgn2\_6/ptodata/2/pubpaa/US08\_PUBCOMB.pep.\*
- 9: /cgn2\_6/ptodata/2/pubpaa/US09A\_PUBCOMB.pep.\*
- 10: /cgn2\_6/ptodata/2/pubpaa/US09B\_PUBCOMB.pep.\*
- 11: /cgn2\_6/ptodata/2/pubpaa/US09C\_PUBCOMB.pep.\*
- 12: /cgn2\_6/ptodata/2/pubpaa/US09\_NEW\_PUB.pep.\*
- 13: /cgn2\_6/ptodata/2/pubpaa/US10A\_PUBCOMB.pep.\*
- 14: /cgn2\_6/ptodata/2/pubpaa/US10B\_PUBCOMB.pep.\*
- 15: /cgn2\_6/ptodata/2/pubpaa/US10C\_PUBCOMB.pep.\*
- 16: /cgn2\_6/ptodata/2/pubpaa/US10\_NEW\_PUB.pep.\*
- 17: /cgn2\_6/ptodata/2/pubpaa/US60\_NEW\_PUB.pep.\*
- 18: /cgn2\_6/ptodata/2/pubpaa/US60\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	357	100.0	263	9	US-09-084-491A-2
2	357	100.0	263	13	US-10-102-704-2
3	357	100.0	263	13	US-10-057-951-2
4	357	100.0	263	14	US-10-210-951-44
5	357	100.0	263	14	US-10-211-884-44
6	146	40.9	527	10	US-09-987-457-18
7	146	40.9	527	10	US-09-987-455-19
8	146	40.9	527	15	US-10-360-101-203
9	146	40.9	562	9	US-09-969-271-7
10	146	40.9	562	9	US-09-974-298-145
11	146	40.9	562	12	US-10-411-037-26
12	146	40.9	562	14	US-10-193-656-8
13	146	40.9	562	14	US-10-443-701-4
14	146	40.9	650	15	US-10-401-077-1
15	141	39.5	655	14	US-10-172-712-28

16	132	37.0	86	15	US-10-233-675A-1	Sequence 1, Appli
17	132	37.0	87	15	US-10-233-675A-10	Sequence 10, Appli
18	132	37.0	322	15	US-10-233-675A-20	Sequence 20, Appli
19	132	37.0	322	15	US-10-233-675A-21	Sequence 21, Appli
20	132	37.0	672	15	US-10-233-675A-15	Sequence 15, Appli
21	132	37.0	674	15	US-10-233-675A-14	Sequence 14, Appli
22	132	37.0	687	15	US-10-233-675A-17	Sequence 17, Appli
23	132	37.0	688	15	US-10-233-675A-18	Sequence 18, Appli
24	132	37.0	689	15	US-10-233-675A-13	Sequence 13, Appli
25	131	36.7	86	15	US-10-233-675A-7	Sequence 7, Appli
26	131	36.7	86	15	US-10-233-675A-22	Sequence 22, Appli
27	131	36.7	86	15	US-10-233-675A-27	Sequence 27, Appli
28	131	36.7	86	15	US-10-233-675A-9	Sequence 9, Appli
29	131	36.7	87	15	US-09-880-503-1	Sequence 1, Appli
30	131	36.7	88	9	US-09-880-503-9	Sequence 9, Appli
31	131	36.7	96	9	US-09-880-503-4	Sequence 4, Appli
32	131	36.7	135	9	US-09-880-503-4	Sequence 12, Appli
33	131	36.7	138	9	US-09-984-186-12	Sequence 12, Appli
34	131	36.7	138	14	US-10-237-667-12	Sequence 12, Appli
35	131	36.7	138	14	US-10-237-708-12	Sequence 12, Appli
36	131	36.7	138	14	US-10-237-866-12	Sequence 12, Appli
37	131	36.7	138	14	US-10-237-871-12	Sequence 12, Appli
38	131	36.7	138	14	US-10-237-624-12	Sequence 12, Appli
39	131	36.7	143	9	US-09-880-503-8	Sequence 8, Appli
40	131	36.7	337	14	US-10-106-698-6266	Sequence 6266, Ap
41	131	36.7	337	15	US-10-264-049-2927	Sequence 2927, Ap
42	131	36.7	403	9	US-09-880-503-6	Sequence 6, Appli
43	131	36.7	411	9	US-09-880-503-3	Sequence 3, Appli
44	131	36.7	411	15	US-10-407-821-2	Sequence 2, Appli
45	131	36.7	431	9	US-09-264-468B-1	Sequence 1, Appli

ALIGNMENTS

RESULT 1  
US-09-084-491A-2  
; Sequence 2, Application US/09084491A  
; Patent No. US20020061576A1  
; GENERAL INFORMATION:  
; APPLICANT: MOORE, PAUL A.  
; APPLICANT: RUBEN, STEVEN M.  
; APPLICANT: EBNER, REINHARD  
; TITLE OF INVENTION: TISSUE PLASMINOGEN ACTIVATOR-LIKE PROTEASE  
; NUMBER OF SEQUENCES: 16  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: HUMAN GENOME SCIENCES, INC.  
; STREET: 9410 KEY WEST AVENUE  
; CITY: ROCKVILLE  
; STATE: MD  
; COUNTRY: USA  
; ZIP: 20850  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/084,491A  
; FILING DATE: 27-MAY-1998  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: BROOKES, ANDERS A.  
; REGISTRATION NUMBER: 36,373  
; REFERENCE/DOCKET NUMBER: PF378  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (301) 309-8504  
; TELEFAX: (301) 309-8439  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 263 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear



```
; MOLECULE TYPE: protein
US-09-084-491A-2

Query Match      100.0%; Score 357; DB 9; Length 263;
Best Local Similarity 100.0%; Pred. No. 1.4e-32;
Matches 60; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CFWDNGHLYREDQTSAPGLRCLNWLDAQSGLASAPVSGAGNHSYCRNPDEDPRGPWCYV 60
    |||||
Db 25 CFWDNGHLYREDQTSAPGLRCLNWLDAQSGLASAPVSGAGNHSYCRNPDEDPRGPWCYV 84

RESULT 2
US-10-102-704-2
; Sequence 2, Application US/10102704
; Publication No. US20020164768A1
; GENERAL INFORMATION:
; APPLICANT: Moore et al.
; TITLE OF INVENTION: Tissue Plasminogen Activator-Like Protein
; FILE REFERENCE: PF378C1
; CURRENT APPLICATION NUMBER: US/10/102,704
; CURRENT FILING DATE: 2002-03-22
; PRIOR APPLICATION NUMBER: 09/084,491
; PRIOR FILING DATE: 1998-05-27
; PRIOR APPLICATION NUMBER: 60/048,000
; PRIOR FILING DATE: 1997-05-28
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 2
; LENGTH: 263
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-102-704-2

Query Match      100.0%; Score 357; DB 13; Length 263;
Best Local Similarity 100.0%; Pred. No. 1.4e-32;
Matches 60; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CFWDNGHLYREDQTSAPGLRCLNWLDAQSGLASAPVSGAGNHSYCRNPDEDPRGPWCYV 60
    |||||
Db 25 CFWDNGHLYREDQTSAPGLRCLNWLDAQSGLASAPVSGAGNHSYCRNPDEDPRGPWCYV 84

RESULT 3
US-10-057-951-2
; Sequence 2, Application US/10057951
; Publication No. US20020177213A1
; GENERAL INFORMATION:
; APPLICANT: Moore et al.
; TITLE OF INVENTION: Tissue Plasminogen Activator-Like Protease
; FILE REFERENCE: PF378P1
; CURRENT APPLICATION NUMBER: US/10/057,951
; CURRENT FILING DATE: 2002-01-29
; PRIOR APPLICATION NUMBER: US 09/411,977
; PRIOR FILING DATE: 1999-10-04
; PRIOR APPLICATION NUMBER: US 09/084,491
; PRIOR FILING DATE: 1998-05-27
; PRIOR APPLICATION NUMBER: US 60/048,000
; PRIOR FILING DATE: 1997-05-28
; NUMBER OF SEQ ID NOS: 31
; SOFTWARE: PatentIn Ver. 3.1
; SEQ ID NO 2
; LENGTH: 263
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-057-951-2

Query Match      100.0%; Score 357; DB 13; Length 263;
Best Local Similarity 100.0%; Pred. No. 1.4e-32;
Matches 60; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CFWDNGHLYREDQTSAPGLRCLNWLDAQSGLASAPVSGAGNHSYCRNPDEDPRGPWCYV 60
    |||||
Db 25 CFWDNGHLYREDQTSAPGLRCLNWLDAQSGLASAPVSGAGNHSYCRNPDEDPRGPWCYV 84

RESULT 4
US-10-210-951-44
; Sequence 44, Application US/10210951
; Publication No. US20030170228A1
; GENERAL INFORMATION:
; APPLICANT: Ashkenazi, Avi J.
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Hillan, Kenneth J.
; APPLICANT: Marsters, Scott A.
; APPLICANT: Pan, James
; APPLICANT: Picti, Robert M.
; APPLICANT: Roy, Margaret Ann
; APPLICANT: Smith, Victoria
; APPLICANT: Stone, Donna M.
; APPLICANT: Watanabe, Colin K.
; APPLICANT: Wood, William I.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE TREATMENT OF TUMOR
; FILE REFERENCE: P2931R1C1
; CURRENT APPLICATION NUMBER: US/10/210,951
; CURRENT FILING DATE: 2002-08-02
; PRIOR APPLICATION NUMBER: 60/014699
; PRIOR FILING DATE: 1996-04-01
; PRIOR APPLICATION NUMBER: 60/026943
; PRIOR FILING DATE: 1996-09-23
; PRIOR APPLICATION NUMBER: 60/059121
; PRIOR FILING DATE: 1997-07-17
; PRIOR APPLICATION NUMBER: 60/059352
; PRIOR FILING DATE: 1997-09-19
; PRIOR APPLICATION NUMBER: 60/062037
; PRIOR FILING DATE: 1997-10-10
; PRIOR APPLICATION NUMBER: 60/063755
; PRIOR FILING DATE: 1997-10-17
; PRIOR APPLICATION NUMBER: 60/063045
; PRIOR FILING DATE: 1997-10-24
; PRIOR APPLICATION NUMBER: 60/063046
; PRIOR FILING DATE: 1997-10-24
; PRIOR APPLICATION NUMBER: 60/066511
; PRIOR FILING DATE: 1997-11-24
; PRIOR APPLICATION NUMBER: 60/066772
; PRIOR FILING DATE: 1997-11-24
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 258
; SEQ ID NO 44
; LENGTH: 263
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-210-951-44

Query Match      100.0%; Score 357; DB 14; Length 263;
Best Local Similarity 100.0%; Pred. No. 1.4e-32;
Matches 60; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CFWDNGHLYREDQTSAPGLRCLNWLDAQSGLASAPVSGAGNHSYCRNPDEDPRGPWCYV 60
    |||||
Db 25 CFWDNGHLYREDQTSAPGLRCLNWLDAQSGLASAPVSGAGNHSYCRNPDEDPRGPWCYV 84

RESULT 5
US-10-211-884-44
; Sequence 44, Application US/10211884
; Publication No. US20030175900A1
; GENERAL INFORMATION:
; APPLICANT: Ashkenazi, Avi J.
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Hillan, Kenneth J.
; APPLICANT: Marsters, Scott A.
```

```

; APPLICANT: Pan,James
; APPLICANT: Pitti,Robert M.
; APPLICANT: Roy,Margaret Ann
; APPLICANT: Smith,Victoria
; APPLICANT: Stone,Donna M.
; APPLICANT: Watanabe,Colin K.
; APPLICANT: Wood,William J.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE TREATMENT OF TUMOR
; FILE REFERENCE: P2931R1C1
; CURRENT APPLICATION NUMBER: US/10/211,884
; CURRENT FILING DATE: 2002-08-02
; PRIOR APPLICATION NUMBER: 60/014699
; PRIOR FILING DATE: 1996-04-01
; PRIOR APPLICATION NUMBER: 60/026943
; PRIOR FILING DATE: 1996-09-23
; PRIOR APPLICATION NUMBER: 60/059121
; PRIOR FILING DATE: 1997-07-17
; PRIOR APPLICATION NUMBER: 60/059352
; PRIOR FILING DATE: 1997-09-19
; PRIOR APPLICATION NUMBER: 60/062037
; PRIOR FILING DATE: 1997-10-10
; PRIOR APPLICATION NUMBER: 60/063755
; PRIOR FILING DATE: 1997-10-17
; PRIOR APPLICATION NUMBER: 60/063045
; PRIOR FILING DATE: 1997-10-24
; PRIOR APPLICATION NUMBER: 60/063046
; PRIOR FILING DATE: 1997-10-24
; PRIOR APPLICATION NUMBER: 60/066511
; PRIOR FILING DATE: 1997-11-24
; PRIOR APPLICATION NUMBER: 60/066772
; PRIOR FILING DATE: 1997-11-24
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 258
; SEQ ID NO 44
; LENGTH: 263
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-211-884-44

Query Match 100.0%; Score 357; DB 14; Length 263;
Best Local Similarity 100.0%; Pred. No. 1.4e-32;
Matches 60; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CFWDNGHLYREDQTSAPGLRCLNWLDAQSGLASAPVSGAGNHSYCRNPDEDPRGPWCYV 60
Db 25 CFWDNGHLYREDQTSAPGLRCLNWLDAQSGLASAPVSGAGNHSYCRNPDEDPRGPWCYV 84

RESULT 6
US-09-987-457-18
; Sequence 18, Application US/09987457
; Publication No. US20030013150A1
; GENERAL INFORMATION:
; APPLICANT: Manosroi, Aranya
; APPLICANT: Manosroi, Jiradej
; APPLICANT: Tayapiwatana, Chatchai
; APPLICANT: Goetz, Friedrich
; APPLICANT: Werner, Rolf-Guenther
; TITLE OF INVENTION: Methods for Large Scale Protein Production in Prokaryotes
; FILE REFERENCE: 0652.2180001
; CURRENT APPLICATION NUMBER: US/09/987,457
; CURRENT FILING DATE: 2001-11-14
; PRIOR APPLICATION NUMBER: 60/268,573
; PRIOR FILING DATE: 2001-02-15
; PRIOR APPLICATION NUMBER: GB 00 27 782.2
; PRIOR FILING DATE: 2000-11-14
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 18
; LENGTH: 527
; TYPE: PRT
; ORGANISM: Homo sapiens (tpa)
US-09-987-457-18

```

```

Query Match 40.9%; Score 146; DB 10; Length 527;
Best Local Similarity 43.5%; Pred. No. 3.1e-08;
Matches 30; Conservative 5; Mismatches 22; Indels 12; Gaps 3;

Qy 1 CFWDNGHLYREDQTSAPGLRCLNWLDAQSGLASAPVSGAGNHSYCRNPDE 51
Db 92 CYEDQGISYRGTSWSTAESGAECTNW--NBSALAQKPYSGRRPDARLGLGNHNYCRNPDR 149

Qy 52 DPRGPWCYV 60
Db 150 DSK-PWCYV 157

RESULT 7
US-09-987-455-19
; Sequence 19, Application US/09987455
; Publication No. US20030049729A1
; GENERAL INFORMATION:
; APPLICANT: Aranya Manosroi
; APPLICANT: Jiradej Manosroi
; APPLICANT: Chatchai Tayapiwatana
; APPLICANT: Friedrich Goetz
; APPLICANT: Rolf-Guenther Werner
; TITLE OF INVENTION: Methods for Large Scale Production of Recombinant
; FILE REFERENCE: 0652.2190001
; CURRENT APPLICATION NUMBER: US/09/987,455
; CURRENT FILING DATE: 2001-11-14
; PRIOR APPLICATION NUMBER: 60/268,574
; PRIOR FILING DATE: 2001-02-15
; PRIOR APPLICATION NUMBER: GB 0027779.8
; PRIOR FILING DATE: 2000-11-14
; NUMBER OF SEQ ID NOS: 25
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 19
; LENGTH: 527
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-987-455-19

Query Match 40.9%; Score 146; DB 10; Length 527;
Best Local Similarity 43.5%; Pred. No. 3.1e-08;
Matches 30; Conservative 5; Mismatches 22; Indels 12; Gaps 3;

Qy 1 CFWDNGHLYREDQTSAPGLRCLNWLDAQSGLASAPVSGAGNHSYCRNPDE 51
Db 92 CYEDQGISYRGTSWSTAESGAECTNW--NBSALAQKPYSGRRPDARLGLGNHNYCRNPDR 149

Qy 52 DPRGPWCYV 60
Db 150 DSK-PWCYV 157

RESULT 8
US-10-360-101-203
; Sequence 203, Application US/10360101
; Publication No. US20040009550A1
; GENERAL INFORMATION:
; APPLICANT: Moll, Gert N.
; APPLICANT: Leenhouts, Cornelis J.
; TITLE OF INVENTION: Export and modification of (poly)peptide in the lantibiotic way
; FILE REFERENCE: 2183-5673
; CURRENT APPLICATION NUMBER: US/10/360,101
; CURRENT FILING DATE: 2003-02-07
; PRIOR APPLICATION NUMBER: EP 02077060.8
; PRIOR FILING DATE: 2002-05-24
; NUMBER OF SEQ ID NOS: 309
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 203
; LENGTH: 527
; TYPE: PRT
; ORGANISM: Artificial Sequence

```

```

RESULT 10
US-09-974-298-145
; Sequence 145, Application US/09974298
; Patent No. US20020156263A1
; GENERAL INFORMATION:
; APPLICANT: Chen, Hwei-Mei
; TITLE OF INVENTION: GENES EXPRESSED IN BREAST CANCER
; FILE REFERENCE: PA-0037 P
; CURRENT APPLICATION NUMBER: US/09/974,298
; CURRENT FILING DATE: 2001-10-04
; PRIOR APPLICATION NUMBER: 60/238,331
; PRIOR FILING DATE: 2000-05-10
; PRIOR OF SEQ ID NOS: 194
; SOFTWARE: PERL Program
; SEQ ID NO 145
; LENGTH: 562
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: Incyte ID No. US20020156263A1 1001470CD1
US-09-974-298-145

```

RESULT 12  
US-10-193-656-8  
; Sequence 8, Application US/10193656  
; Publication No. US20030096733A1  
; GENERAL INFORMATION:  
; APPLICANT: NY, Tor  
; APPLICANT: HOLMADHL, Rikard

```
; APPLICANT: LI, Jinan
; TITLE OF INVENTION: NOVEL DRUG TARGETS FOR ARTHRITIS
; FILE REFERENCE: 3810/10577-US3
; CURRENT APPLICATION NUMBER: US/10/193,656
; PRIOR FILING DATE: 2002-07-10
; PRIOR APPLICATION NUMBER: US 60/304,461
; PRIOR FILING DATE: 2001-07-10
; PRIOR APPLICATION NUMBER: US 60/304,490
; PRIOR FILING DATE: 2001-07-10
; PRIOR APPLICATION NUMBER: US 60/305,182
; PRIOR FILING DATE: 2001-07-13
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: Patent in version 3.1
; SEQ ID NO 8
; LENGTH: 562
; TYPE: PRT
; ORGANISM: Homo sapiens
; PUBLICATION INFORMATION:
; DATABASE ACCESSION NUMBER: GenBank / P00750
; DATABASE ENTRY DATE: 1986-07-21
; RELEVANT RESIDUES: (1)..(562)
US-10-193-656-8

Query Match          40.9%; Score 146; DB 14; Length 562;
Best Local Similarity 43.5%; Pred. No. 3.3e-08;
Matches 30; Conservative 5; Mismatches 22; Indels 12; Gaps 3;

QY 1 CFWDNGHLYREDQTSAPGLRCLNWLDAOSGLASAPVS-----GAGNHSYCRNPDE 51
   |||||
Db 127 CYEDQGISYRGTSWTAESGAECTNW--NSSALAQKPYSGRRPDALRLGLGNHNYCRNPDR 184
   |||||

QY 52 DPGPWCYV 60
   |||||
Db 185 DSK-PWCYV 192
   |||||

RESULT 13
US-10-443-701-4
; Sequence 4, Application US/10443701
; Publication No. US20030195016A1
; GENERAL INFORMATION:
; APPLICANT: Xu, Yuan
; TITLE OF INVENTION: REVERSE-PHASE HPLC ASSAY FOR PLASMINOGEN ACTIVATORS
; FILE REFERENCE: P1789R1
; CURRENT APPLICATION NUMBER: US/10/443,701
; CURRENT FILING DATE: 2003-05-21
; PRIOR APPLICATION NUMBER: US/03/703,695
; PRIOR FILING DATE: 2000-11-01
; PRIOR APPLICATION NUMBER: US 60/163,607
; PRIOR FILING DATE: 1999-11-04
; NUMBER OF SEQ ID NOS: 4
; SEQ ID NO 4
; LENGTH: 562
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-443-701-4

Query Match          40.9%; Score 146; DB 14; Length 562;
Best Local Similarity 43.5%; Pred. No. 3.3e-08;
Matches 30; Conservative 5; Mismatches 22; Indels 12; Gaps 3;

QY 1 CFWDNGHLYREDQTSAPGLRCLNWLDAOSGLASAPVS-----GAGNHSYCRNPDE 51
   |||||
Db 127 CYEDQGISYRGTSWTAESGAECTNW--NSSALAQKPYSGRRPDALRLGLGNHNYCRNPDR 184
   |||||

QY 52 DPGPWCYV 60
   |||||
Db 185 DSK-PWCYV 192
   |||||

; APPLICANT: LI, Jinan
; TITLE OF INVENTION: NOVEL DRUG TARGETS FOR ARTHRITIS
; FILE REFERENCE: 3810/10577-US3
; CURRENT APPLICATION NUMBER: US/10/193,656
; PRIOR FILING DATE: 2002-07-10
; PRIOR APPLICATION NUMBER: US 60/304,461
; PRIOR FILING DATE: 2001-07-10
; PRIOR APPLICATION NUMBER: US 60/304,490
; PRIOR FILING DATE: 2001-07-10
; PRIOR APPLICATION NUMBER: US 60/305,182
; PRIOR FILING DATE: 2001-07-13
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: Patent in version 3.1
; SEQ ID NO 8
; LENGTH: 562
; TYPE: PRT
; ORGANISM: Homo sapiens
; PUBLICATION INFORMATION:
; DATABASE ACCESSION NUMBER: GenBank / P00750
; DATABASE ENTRY DATE: 1986-07-21
; RELEVANT RESIDUES: (1)..(562)
US-10-193-656-8

Query Match          40.9%; Score 146; DB 14; Length 562;
Best Local Similarity 43.5%; Pred. No. 3.3e-08;
Matches 30; Conservative 5; Mismatches 22; Indels 12; Gaps 3;

QY 1 CFWDNGHLYREDQTSAPGLRCLNWLDAOSGLASAPVS-----GAGNHSYCRNPDE 51
   |||||
Db 127 CYEDQGISYRGTSWTAESGAECTNW--NSSALAQKPYSGRRPDALRLGLGNHNYCRNPDR 184
   |||||

QY 52 DPGPWCYV 60
   |||||
Db 185 DSK-PWCYV 192
   |||||

; APPLICANT: Hung, Paul Porwen
; APPLICANT: Wu, Bryan T. H.
; TITLE OF INVENTION: HUMAN TISSUE UROKINASE TYPE PLASMINOGEN
; TITLE OF INVENTION: ACTIVATOR PRODUCTION
; FILE REFERENCE: 12133-006001
; CURRENT APPLICATION NUMBER: US/10/401,077
; CURRENT FILING DATE: 2003-03-27
; PRIOR APPLICATION NUMBER: US 60/371,013
; PRIOR FILING DATE: 2002-04-09
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 650
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-401-077-1

Query Match          40.9%; Score 146; DB 15; Length 650;
Best Local Similarity 43.5%; Pred. No. 3.8e-08;
Matches 30; Conservative 5; Mismatches 22; Indels 12; Gaps 3;

QY 1 CFWDNGHLYREDQTSAPGLRCLNWLDAOSGLASAPVS-----GAGNHSYCRNPDE 51
   |||||
Db 215 CYEDQGISYRGTSWTAESGAECTNW--NSSALAQKPYSGRRPDALRLGLGNHNYCRNPDR 272
   |||||

QY 52 DPGPWCYV 60
   |||||
Db 273 DSK-PWCYV 280
   |||||

RESULT 15
US-10-172-712-28
; Sequence 28, Application US/10172712
; Publication No. US20030125232A1
; GENERAL INFORMATION:
; APPLICANT: GRIFFIN, JOHN H.
; APPLICANT: GALE, ANDREW J.
; APPLICANT: GETZOFF, ELIZABETH D.
; APPLICANT: PELLEQUER, JEAN-LUC
; TITLE OF INVENTION: STABILIZED PROTEINS WITH ENGINEERED DISULFIDE BONDS
; FILE REFERENCE: 4198-4001US1
; CURRENT APPLICATION NUMBER: US/10/172,712
; CURRENT FILING DATE: 2002-09-30
; PRIOR APPLICATION NUMBER: 60/298,578
; PRIOR FILING DATE: 2001-06-14
; NUMBER OF SEQ ID NOS: 32
; SOFTWARE: Patent in Ver. 2.1
; SEQ ID NO 28
; LENGTH: 655
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-172-712-28

Query Match          39.5%; Score 141; DB 14; Length 655;
Best Local Similarity 45.6%; Pred. No. 1.4e-07;
Matches 31; Conservative 6; Mismatches 21; Indels 10; Gaps 3;

QY 1 CFWDNGHLYREDQTSAPGLRCLNWLDAOSGLASAPVS-----GAGNHSYCRNPDE 52
   |||||
Db 286 CFLGNGTGYRGVASTSASGLSCLAWNSDLLYQELHVDV-VGAAALLGLGPHAYCRNPDPD 344
   |||||

QY 53 PRGPWCYV 60
   |||||
Db 345 ER-PWCYV 351
   |||||

Search completed: March 18, 2004, 13:24:46
Job time : 11.9017 secs
```

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 17, 2004, 07:00:39 ; Search time 4.4145 Seconds  
(without alignments)  
697.420 Million cell updates/sec

Title: US-10-057-951-2\_COPY\_25\_84

Perfect score: 357  
Sequence: 1 CFWDNGHLYREDQTSAPGL.....GNHSYCRNPDEDRGPWCYV 60

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued Patents AA:\*  
1: /cn2\_6/prodata/2/aaa/5A-COMB.pep:\*  
2: /cn2\_6/prodata/2/aaa/5B-COMB.pep:\*  
3: /cn2\_6/prodata/2/aaa/6A-COMB.pep:\*  
4: /cn2\_6/prodata/2/aaa/6B-COMB.pep:\*  
5: /cn2\_6/prodata/2/aaa/6C-COMB.pep:\*  
6: /cn2\_6/prodata/2/aaa/6D-COMB.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	357	100.0	263	4	US-09-411-977-2
2	150	42.0	472	2	US-08-811-949-63
3	146	40.9	83	2	US-08-811-949-2
4	146	40.9	437	2	US-08-811-949-49
5	146	40.9	437	2	US-08-811-949-51
6	146	40.9	437	2	US-08-811-949-55
7	146	40.9	437	2	US-08-811-949-57
8	146	40.9	527	1	US-07-609-510B-16
9	146	40.9	527	2	US-08-811-949-39
10	146	40.9	527	5	PCT-US91-01025A-2
11	146	40.9	527	6	5185259-8
12	146	40.9	527	6	5203013-1
13	146	40.9	546	6	5200340-6
14	146	40.9	562	2	US-08-811-949-43
15	146	40.9	562	2	US-08-560-098A-50
16	146	40.9	562	4	US-08-883-795A-38
17	146	40.9	562	4	US-09-703-695A-4
18	146	40.9	562	6	5185259-3
19	146	40.9	562	6	5200340-2
20	146	40.9	562	6	5344773-2
21	141	39.5	655	1	US-08-148-910-12
22	141	39.5	655	1	US-08-448-937A-12
23	139	38.5	356	1	US-08-427-640-8
24	134	37.5	477	2	US-08-560-098A-51
25	131	36.7	89	4	US-09-101-272G-62
26	131	36.7	138	2	US-08-737-689-12
27	131	36.7	138	4	US-09-984-186-12

28	131	36.7	194	4	US-09-101-272G-80	Sequence 80, Appl
29	131	36.7	200	4	US-09-101-272G-73	Sequence 73, Appl
30	131	36.7	201	4	US-09-101-272G-96	Sequence 96, Appl
31	131	36.7	208	4	US-09-101-272G-98	Sequence 98, Appl
32	131	36.7	365	1	US-08-093-741-83	Sequence 83, Appl
33	131	36.7	365	1	US-08-720-012-83	Sequence 83, Appl
34	131	36.7	393	2	US-08-560-098A-44	Sequence 44, Appl
35	131	36.7	393	3	US-08-967-024C-24	Sequence 24, Appl
36	131	36.7	393	3	US-08-967-024C-25	Sequence 25, Appl
37	131	36.7	411	1	US-08-087-163-1	Sequence 1, Appl
38	131	36.7	411	1	US-08-286-748B-18	Sequence 18, Appl
39	131	36.7	411	1	US-08-153-799-18	Sequence 18, Appl
40	131	36.7	411	2	US-08-560-098A-48	Sequence 48, Appl
41	131	36.7	411	3	US-09-181-816-1	Sequence 1, Appl
42	131	36.7	411	3	US-09-403-736-2	Sequence 3, Appl
43	131	36.7	430	1	US-07-942-157A-3	Sequence 3, Appl
44	131	36.7	430	6	5219569-2	Patent No. 5219569
45	131	36.7	431	4	US-09-101-272G-1	Sequence 1, Appl

ALIGNMENTS

RESULT 1  
US-09-411-977-2  
; Sequence 2, Application US/09411977  
; Patent No. 6372473

GENERAL INFORMATION:  
; APPLICANT: Moore, Paul A.  
; APPLICANT: Ruben, Steven M.  
; APPLICANT: Ebner, Reinhard  
; TITLE OF INVENTION: Tissue Plasminogen Activator-Like Protease  
; FILE REFERENCE: PF378P1  
; CURRENT APPLICATION NUMBER: US/09/411,977  
; EARLIER FILING DATE: 1989-10-04  
; EARLIER APPLICATION NUMBER: 09/084,491  
; EARLIER FILING DATE: 1998-05-27  
; EARLIER APPLICATION NUMBER: 60/048,000  
; EARLIER FILING DATE: 1997-05-28  
; NUMBER OF SEQ ID NOS: 30  
; SOFTWARE: Patent in Ver. 2.0  
; SEQ ID NO 2  
; LENGTH: 263  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-411-977-2

Query Match 100.0%; Score 357; DB 4; Length 263;  
Best Local Similarity 100.0%; Pred. No. 1.4e-35;  
Matches 60; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CFWDNGHLYREDQTSAPGLRCLNWLDAQSLASAPYSGAGNHSYCRNPDEDRGPWCYV 60  
DB 25 CFWDNGHLYREDQTSAPGLRCLNWLDAQSLASAPYSGAGNHSYCRNPDEDRGPWCYV 84

RESULT 2  
US-08-811-949-63  
; Sequence 63, Application US/08811949  
; Patent No. 5840533

GENERAL INFORMATION:  
; APPLICANT: NIWA, MINEO  
; APPLICANT: SAITO, YOSHIMASA  
; APPLICANT: SASAKI, HITOSHI  
; APPLICANT: HAYASHI, MASAKO  
; APPLICANT: NOTANI, JOUJI  
; APPLICANT: KOBAYASHI, MASAKAZU  
; TITLE OF INVENTION: TISSUE PLASMINOGEN ACTIVATOR  
; NUMBER OF SEQUENCES: 67  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MAIER & NEUSTADT,  
; ADDRESS: P.C.  
; STREET: 1755 S. JEFFERSON DAVIS HIGHWAY, SUITE 400

```

; CITY: ARLINGTON
; STATE: VA USA
; COUNTRY: USA
; ZIP: 22202
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/811,949
; FILING DATE: 05-MAR-1997
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: OBLON, NORMAN F.
; REGISTRATION NUMBER: 24,618
; REFERENCE/DOCKET NUMBER: 18-966-0
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 703-413-3000
; TELEFAX: 703-413-2220
; INFORMATION FOR SEQ ID NO: 63:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 472 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: Protein
;
; US-08-811-949-63
;
; Query Match 42.0%; Score 150; DB 2; Length 472;
; Best Local Similarity 44.9%; Pred. No. 3.8e-10;
; Matches 31; Conservative 4; Mismatches 22; Indels 12; Gaps 3;
;
; QY 1 CFWDNHLYREDOTSPAPGLRCLNWLDAQSLASAPVS-----GAGNHSYCRNPDE 51
; Db 37 CYEDQGISYRGVTWTAESGAECTNW--NSSALAQKPYSGRRPDPIRLGLGNHNYCRNPDR 94
;
; QY 52 DPRGPWCYV 60
; Db 95 DSK-PWCYV 102
;
; RESULT 3
; US-08-811-949-2
; Sequence 2, Application US/08811949
; Patent No. 5840533
; GENERAL INFORMATION:
; APPLICANT: NIWA, MINEO
; APPLICANT: SAITO, YOSHIMASA
; APPLICANT: SASAKI, HITOSHI
; APPLICANT: HAYASHI, MASAKO
; APPLICANT: NOTANI, JOUJI
; APPLICANT: KOBAYASHI, MASAKAZU
; TITLE OF INVENTION: TISSUE PLASMINOGEN ACTIVATOR
; NUMBER OF SEQUENCES: 67
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MAIER & NEUSTADT,
; ADDRESSEE: P.C.
; STREET: 1755 S. JEFFERSON DAVIS HIGHWAY, SUITE 400
; CITY: ARLINGTON
; STATE: VA
; COUNTRY: USA
; ZIP: 22202
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/811,949
; FILING DATE: 05-MAR-1997
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: OBLON, NORMAN F.
; REGISTRATION NUMBER: 24,618
; REFERENCE/DOCKET NUMBER: 18-966-0
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 703-413-3000
; TELEFAX: 703-413-2220
; INFORMATION FOR SEQ ID NO: 49:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 437 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
;
; US-08-811-949-49
;
; Query Match 40.9%; Score 146; DB 2; Length 437;
; Best Local Similarity 43.5%; Pred. No. 1.5e-10;
; Matches 30; Conservative 5; Mismatches 22; Indels 12; Gaps 3;
;
; QY 1 CFWDNHLYREDOTSPAPGLRCLNWLDAQSLASAPVS-----GAGNHSYCRNPDE 51
; Db 1 CYEDQGISYRGVTWTAESGAECTNW--NSSALAQKPYSGRRPDPIRLGLGNHNYCRNPDR 58
;
; QY 52 DPRGPWCYV 60
; Db 59 DSK-PWCYV 66
;
; RESULT 4
; US-08-811-949-49
; Sequence 49, Application US/08811949
; Patent No. 5840533
; GENERAL INFORMATION:
; APPLICANT: NIWA, MINEO
; APPLICANT: SAITO, YOSHIMASA
; APPLICANT: SASAKI, HITOSHI
; APPLICANT: HAYASHI, MASAKO
; APPLICANT: NOTANI, JOUJI
; APPLICANT: KOBAYASHI, MASAKAZU
; TITLE OF INVENTION: TISSUE PLASMINOGEN ACTIVATOR
; NUMBER OF SEQUENCES: 67
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MAIER & NEUSTADT,
; ADDRESSEE: P.C.
; STREET: 1755 S. JEFFERSON DAVIS HIGHWAY, SUITE 400
; CITY: ARLINGTON
; STATE: VA
; COUNTRY: USA
; ZIP: 22202
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/811,949
; FILING DATE: 05-MAR-1997
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: OBLON, NORMAN F.
; REGISTRATION NUMBER: 24,618
; REFERENCE/DOCKET NUMBER: 18-966-0
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 703-413-3000
; TELEFAX: 703-413-2220
; INFORMATION FOR SEQ ID NO: 49:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 437 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
;
; US-08-811-949-49
;
; Query Match 40.9%; Score 146; DB 2; Length 437;
```

```

; REGISTRATION NUMBER: 24,618
; REFERENCE/DOCKET NUMBER: 18-966-0
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 703-413-3000
; TELEFAX: 703-413-2220
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 83 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
;
; US-08-811-949-2
;
; Query Match 40.9%; Score 146; DB 2; Length 83;
; Best Local Similarity 43.5%; Pred. No. 1.5e-10;
; Matches 30; Conservative 5; Mismatches 22; Indels 12; Gaps 3;
;
; QY 1 CFWDNHLYREDOTSPAPGLRCLNWLDAQSLASAPVS-----GAGNHSYCRNPDE 51
; Db 1 CYEDQGISYRGVTWTAESGAECTNW--NSSALAQKPYSGRRPDPIRLGLGNHNYCRNPDR 58
;
; QY 52 DPRGPWCYV 60
; Db 59 DSK-PWCYV 66
;
; RESULT 4
; US-08-811-949-49
; Sequence 49, Application US/08811949
; Patent No. 5840533
; GENERAL INFORMATION:
; APPLICANT: NIWA, MINEO
; APPLICANT: SAITO, YOSHIMASA
; APPLICANT: SASAKI, HITOSHI
; APPLICANT: HAYASHI, MASAKO
; APPLICANT: NOTANI, JOUJI
; APPLICANT: KOBAYASHI, MASAKAZU
; TITLE OF INVENTION: TISSUE PLASMINOGEN ACTIVATOR
; NUMBER OF SEQUENCES: 67
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MAIER & NEUSTADT,
; ADDRESSEE: P.C.
; STREET: 1755 S. JEFFERSON DAVIS HIGHWAY, SUITE 400
; CITY: ARLINGTON
; STATE: VA
; COUNTRY: USA
; ZIP: 22202
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/811,949
; FILING DATE: 05-MAR-1997
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: OBLON, NORMAN F.
; REGISTRATION NUMBER: 24,618
; REFERENCE/DOCKET NUMBER: 18-966-0
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 703-413-3000
; TELEFAX: 703-413-2220
; INFORMATION FOR SEQ ID NO: 49:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 437 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
;
; US-08-811-949-49
;
; Query Match 40.9%; Score 146; DB 2; Length 437;
```

```
Best Local Similarity 43.5%; Pred. No. 1.le-09;
Matches 30; Conservative 5; Mismatches 22; Indels 12; Gaps 3;

Qy 1 CFWDNGHLYREDQTSAPGLRCLNWLDAQSLASAPVS-----GAGNHSYCRNPDE 51
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 2 CYEDQGISYRGTSWTAESGAECTNW--NSSALAQKPYSGRRPDARLGLGNHNYCRNPDR 59
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 52 DPRGPWCYV 60
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 60 DSK-PWCYV 67

RESULT 5
US-08-811-949-51
; Sequence 51, Application US/08811949
; Patent No. 5840533
; GENERAL INFORMATION:
; APPLICANT: NIWA, MINEO
; APPLICANT: SAITO, YOSHIMASA
; APPLICANT: SASAKI, HITOSHI
; APPLICANT: HAYASHI, MASAKO
; APPLICANT: NOTANI, JOUJI
; APPLICANT: KOBAYASHI, MASAKAZU
; TITLE OF INVENTION: TISSUE PLASMINOGEN ACTIVATOR
; NUMBER OF SEQUENCES: 67
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MAIER & NEUSTADT,
; STREET: 1755 S. JEFFERSON DAVIS HIGHWAY, SUITE 400
; CITY: ARLINGTON
; STATE: VA
; COUNTRY: USA
; ZIP: 22202
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/811,949
; FILING DATE: 05-MAR-1997
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: OBLON, NORMAN F.
; REGISTRATION NUMBER: 24,618
; REFERENCE/DOCKET NUMBER: 18-966-0
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 703-413-3000
; TELEFAX: 703-413-2220
; INFORMATION FOR SEQ ID NO: 51:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 437 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-811-949-51

Query Match 40.9%; Score 146; DB 2; Length 437;
Best Local Similarity 43.5%; Pred. No. 1.le-09;
Matches 30; Conservative 5; Mismatches 22; Indels 12; Gaps 3;

Qy 1 CFWDNGHLYREDQTSAPGLRCLNWLDAQSLASAPVS-----GAGNHSYCRNPDE 51
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 2 CYEDQGISYRGTSWTAESGAECTNW--NSSALAQKPYSGRRPDARLGLGNHNYCRNPDR 59
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 52 DPRGPWCYV 60
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 60 DSK-PWCYV 67

RESULT 6
US-08-811-949-55
; Sequence 55, Application US/08811949
```

```
Patent No. 5840533
GENERAL INFORMATION:
APPLICANT: NIWA, MINEO
APPLICANT: SAITO, YOSHIMASA
APPLICANT: SASAKI, HITOSHI
APPLICANT: HAYASHI, MASAKO
APPLICANT: NOTANI, JOUJI
APPLICANT: KOBAYASHI, MASAKAZU
TITLE OF INVENTION: TISSUE PLASMINOGEN ACTIVATOR
NUMBER OF SEQUENCES: 67
CORRESPONDENCE ADDRESS:
ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MAIER & NEUSTADT,
STREET: 1755 S. JEFFERSON DAVIS HIGHWAY, SUITE 400
CITY: ARLINGTON
STATE: VA
COUNTRY: USA
ZIP: 22202
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/811,949
FILING DATE: 05-MAR-1997
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: OBLON, NORMAN F.
REGISTRATION NUMBER: 24,618
REFERENCE/DOCKET NUMBER: 18-966-0
TELECOMMUNICATION INFORMATION:
TELEPHONE: 703-413-3000
TELEFAX: 703-413-2220
INFORMATION FOR SEQ ID NO: 55:
SEQUENCE CHARACTERISTICS:
LENGTH: 437 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-811-949-55

Query Match 40.9%; Score 146; DB 2; Length 437;
Best Local Similarity 43.5%; Pred. No. 1.le-09;
Matches 30; Conservative 5; Mismatches 22; Indels 12; Gaps 3;

Qy 1 CFWDNGHLYREDQTSAPGLRCLNWLDAQSLASAPVS-----GAGNHSYCRNPDE 51
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 2 CYEDQGISYRGTSWTAESGAECTNW--NSSALAQKPYSGRRPDARLGLGNHNYCRNPDR 59
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 52 DPRGPWCYV 60
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 60 DSK-PWCYV 67

RESULT 7
US-08-811-949-57
; Sequence 57, Application US/08811949
; Patent No. 5840533
; GENERAL INFORMATION:
; APPLICANT: NIWA, MINEO
; APPLICANT: SAITO, YOSHIMASA
; APPLICANT: SASAKI, HITOSHI
; APPLICANT: HAYASHI, MASAKO
; APPLICANT: NOTANI, JOUJI
; APPLICANT: KOBAYASHI, MASAKAZU
; TITLE OF INVENTION: TISSUE PLASMINOGEN ACTIVATOR
; NUMBER OF SEQUENCES: 67
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MAIER & NEUSTADT,
; STREET: 1755 S. JEFFERSON DAVIS HIGHWAY, SUITE 400
; CITY: ARLINGTON
```

Thu Mar 18 15:28:05 2004

```

; STATE: VA
; COUNTRY: USA
; ZIP: 22202
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/811,949
; FILING DATE: 05-MAR-1997
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: OBLON, NORMAN F.
; REGISTRATION NUMBER: 24,618
; REFERENCE/DOCKET NUMBER: 18-956-0
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 703-413-3000
; TELEFAX: 703-413-2220
; INFORMATION FOR SEQ ID NO: 57:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 437 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-811-949-57

```

Qy	1	CFWDNGHLYREDQTSFAPGLRCIANWLDQAQGLASAPVS-----GAGNSHSYCRNPDE	51
		:                  :                  :                  :	
Db	2	CYEDQGYSYRGTTWSTAESGAECTNW--NSSALAAQKPYSGRRPDAIRLGLGNHNYCRNPDR	59
		:                  :                  :                  :	
Qy	52	DPRGFCWCYV	60
		:                  :                  :                  :	
Db	60	DSK-PWCYV	67

RESULT 8  
US-07-609-510B-16  
; Sequence 16, Application US/07609510B  
; Patent No. 5326700  
; GENERAL INFORMATION:  
; APPLICANT: Berg et al.  
; TITLE OF INVENTION: Method for Altering Post-Translational Processing of Tissue Protein  
; NUMBER OF SEQUENCES: 16  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Eli Lilly and Company  
; STREET: Lilly Corporate Center  
; CITY: Indianapolis  
; STATE: IN.  
; COUNTRY: U.S.A.  
; ZIP: 46285  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette, 3.50 inch, 1.0 Mb storage  
; COMPUTER: Macintosh  
; OPERATING SYSTEM: Macintosh  
; SOFTWARE: Microsoft Word  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/07/609,510B  
; FILING DATE: 19901106  
; CLASSIFICATION: 435  
; INFORMATION FOR SEQ ID NO: 16:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 527 amino acids  
; TYPE: AMINO ACID  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-07-609-510B-16

```

Query Match          40.98; Score 146; DB 1; Length 527;
Best Local Similarity 43.58; Pred No. 1.3e-09;
Matches 30; Conservative 5; Mismatches 22; Indels 12; Gaps 3;

QY 1 CFWNHGHLTYREDQTSFAPGURCLNWLDAQSLASAPVS-----GAGNHSYCRNPDE 51
DB 92 CYEDQGISYRGTWSTAESGAECTNW--NSSALAAQKPYSGRRPDPAIRLGLGNHNYCRNPDR 149
      : : : : :
QY 52 DPRGPCVCYV 60
DB 150 DSK-PWCYV 157
      : : : : :

```

```

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/811,949
FILING DATE: 05-MAR-1997
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: OBOLON, NORMAN F.
REGISTRATION NUMBER: 24,618
REFERENCE/DOCKET NUMBER: 18-966-0
TELECOMMUNICATION INFORMATION:
TELEPHONE: 703-413-3000
TELEFAX: 703-413-2220
INFORMATION FOR SEQ ID NO: 39:
SEQUENCE CHARACTERISTICS:
LENGTH: 527 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-811-949-39

```

```

Query Match      40.9%; Score 146; DB 2; Length 527;
Best Local Similarity 43.5%; Pred. No. 1.3e-07;
Matches 30; Conservative 5; Mismatches 22; Indels 12; Gaps 3;

QY 1 CFWNGXHYREDQTSFAPGLKLNWLDQAQSLGAPVS-----GAGNHSYCRNPDE 51
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db CYEDGIGSYRGTWSTAEGSAECTNW--NSSALQKPSYGRPRDAIRLGLGNHNYCRNPDR 149
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 52 DPRGPWCYV 60
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 150 DSK-PWCYV 157
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

RESULT 10  
PCT-US91-01025A-2



```

;SEQ ID NO:8;
; LENGTH: 527
; 5185259-8

Query Match 40.9%; Score 146; DB 6; Length 527;
Best Local Similarity 43.5%; Pred. No. 1.3e-09;
Matches 30; Conservative 5; Mismatches 22; Indels 12; Gaps 3;

QY 1 CFWDNHLYREDQTSAPAGRLCLNWLDAOSGLASAPVS-----GAGNHSYCRNPDE 51
DB 92 CYEDQGISYRGWTWSTAESGAECTNW--NSSALLAQKPYSGRRPDAIRLGLGNHNYCRNPDR 149
QY 52 DPRGPWCYV 60
DB 150 DSK-PWCYV 157

RESULT 12
5520913-1
; Patent No. 5520913
; APPLICANT: ANDERSON, STEPHEN; BENNETT, WILLIAM F.; BOTSTEIN.
; DAVID; HIGGINS, DEBORAH L.; PAONI, NICHOLAS F.; ZOLLER, MARK J.
; TITLE OF INVENTION: TISSUE PLASMINOGEN ACTIVATOR HAVING
; ZYMOGENIC PROPERTIES
; NUMBER OF SEQUENCES: 35
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/88,451
; FILING DATE: 06-JUL-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 770,510
; FILING DATE: 03-OCT-1991
; APPLICATION NUMBER: 384,608
; FILING DATE: 24-JUL-1989
; APPLICATION NUMBER: 240,856
; FILING DATE: 02-SEP-1988
; SEQ ID NO:1:
; LENGTH: 527
; 5520913-1

Query Match 40.9%; Score 146; DB 6; Length 527;
Best Local Similarity 43.5%; Pred. No. 1.3e-09;
Matches 30; Conservative 5; Mismatches 22; Indels 12; Gaps 3;

QY 1 CFWDNHLYREDQTSAPAGRLCLNWLDAOSGLASAPVS-----GAGNHSYCRNPDE 51
DB 92 CYEDQGISYRGWTWSTAESGAECTNW--NSSALLAQKPYSGRRPDAIRLGLGNHNYCRNPDR 149
QY 52 DPRGPWCYV 60
DB 150 DSK-PWCYV 157

RESULT 13
5200340-6
; Patent No. 5200340
; APPLICANT: POSTER, DONALD C.; MULVIHILL, EILEEN R.; O'HARA,
; PATRICK J.; PINGEL, KURT; YOSHITAKE, SHINJI
; TITLE OF INVENTION: THROMBIN-ACTIVATED TISSUE PLASMINOGEN
; ACTIVATORS
; NUMBER OF SEQUENCES: 34
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/53,412
; FILING DATE: 22-MAY-1987
; SEQ ID NO:6:
; LENGTH: 546
; 5200340-6

Query Match 40.9%; Score 146; DB 6; Length 546;
Best Local Similarity 43.8%; Pred. No. 1.4e-09;
Matches 30; Conservative 5; Mismatches 22; Indels 12; Gaps 3;

QY 1 CFWDNHLYREDQTSAPAGRLCLNWLDAOSGLASAPVS-----GAGNHSYCRNPDE 51

```

Db 127 CYEDQGISYRGWTSAESGAECTNW--NSSALAQKPYSGRRPDARLGLGNHNYCRNPDR 184

QY 52 DPRGPWCYV 60  
| : |||||  
Db 185 DSK-PWCYV 192

## RESULT 14

US-08-811-949-43  
; Sequence 43, Application US/08911949  
; Patent No. 5840533  
; GENERAL INFORMATION:  
; APPLICANT: NIWA, MINEO  
; APPLICANT: SAITO, YOSHIMASA  
; APPLICANT: SASAKI, HITOSHI  
; APPLICANT: HAYASHI, MASAKO  
; APPLICANT: NOTANI, JOUJI  
; APPLICANT: KOBAYASHI, MASAKAZU  
; TITLE OF INVENTION: TISSUE PLASMINOGEN ACTIVATOR  
; NUMBER OF SEQUENCES: 67  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: OBLON, SPIVAK, MCLELLAND, MAIER & NEUSTADT,  
; ADDRESSEE: P.C.  
; STREET: 1755 S. JEFFERSON DAVIS HIGHWAY, SUITE 400  
; CITY: ARLINGTON  
; STATE: VA  
; COUNTRY: USA  
; ZIP: 22202  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/811,949  
; FILING DATE: 05-MAR-1997  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: OBLON, NORMAN F.  
; REGISTRATION NUMBER: 24,618  
; REFERENCE/DOCKET NUMBER: 18-966-0  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 703-413-3000  
; TELEFAX: 703-413-2220  
; INFORMATION FOR SEQ ID NO: 43:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 562 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; US-08-811-949-43

Query Match 40.9%; Score 146; DB 2; Length 562;  
Best Local Similarity 43.5%; Pred. No. 1.4e-09;  
Matches 30; Conservative 5; Mismatches 22; Indels 12; Gaps 3;

QY 1 CFWDNGHLYREDQTSAPAGRLCLNWLDAQSGLASAPVS-----GAGNHSYCRNPDE 51  
| : |||||  
Db 127 CYEDQGISYRGWTSAESGAECTNW--NSSALAQKPYSGRRPDARLGLGNHNYCRNPDR 184

QY 52 DPRGPWCYV 60  
| : |||||  
Db 185 DSK-PWCYV 192

## RESULT 15

US-08-560-098A-50  
; Sequence 50, Application US/08560098A  
; Patent No. 5976841  
; GENERAL INFORMATION:  
; APPLICANT: WENNDT, Stephan  
; APPLICANT: HEINZEL-WIELAND, Regina  
; APPLICANT: STEFFENS, Gerd Josef

; TITLE OF INVENTION: Proteins having Fibrinolytic and  
; TITLE OF INVENTION: Coagulation-inhibiting Properties  
; NUMBER OF SEQUENCES: 60  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Evenson, McKeown, Edwards & Lenahan  
; STREET: 1200 G Street, N.W., Suite 700  
; CITY: Washington  
; STATE: DC  
; COUNTRY: USA  
; ZIP: 20005  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.30 (EPO)  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/560,098A  
; FILING DATE: 17-NOV-1995  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: P 44 40 892.7  
; FILING DATE: 17-NOV-1994  
; ATTORNEY/AGENT INFORMATION:  
; NAME: EVANS, Joseph D.  
; REGISTRATION NUMBER: 26,269  
; REFERENCE/DOCKET NUMBER: 148/42448  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (202) 628-8800  
; TELEFAX: (202) 628-8844  
; INFORMATION FOR SEQ ID NO: 50:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 562 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; US-08-560-098A-50

Query Match 40.9%; Score 146; DB 2; Length 562;  
Best Local Similarity 43.5%; Pred. No. 1.4e-09;  
Matches 30; Conservative 5; Mismatches 22; Indels 12; Gaps 3;

QY 1 CFWDNGHLYREDQTSAPAGRLCLNWLDAQSGLASAPVS-----GAGNHSYCRNPDE 51  
| : |||||  
Db 127 CYEDQGISYRGWTSAESGAECTNW--NSSALAQKPYSGRRPDARLGLGNHNYCRNPDR 184

QY 52 DPRGPWCYV 60  
| : |||||  
Db 185 DSK-PWCYV 192

Search completed: March 17, 2004, 07:09:07  
Job time : 4.44145 secs



A;Gene: CESP:F45E12.2  
A;Accession: F45E12.2  
A;Molecule type: DNA  
A;Residues: 55/1; 152/3; 392/2; 650/3; 691/3; 731/3  
C;Superfamily: transcription initiation factor IIB 90K chain; transcription initiation

Query Match 9.6%; Score 87.5; DB 2; Length 759;  
Best Local Similarity 30.4%; Pred. No. 4.2;  
Matches 28; Conservative 17; Mismatches 32; Indels 15; Gaps 5;

QY 1 SGEAGVPRKPCEDL--RCPETTSQALPAFTTEIQEASEGPADEV-----QVFAPANAL 53  
DB 577 ASESTIQKLSIFDLTEECSETSKNSPKVNLKVESAS--PSTSEVSIEHKFPVP--- 630

QY 54 PARSAAAVQPVIGISQVRVMSKEKDLGTL 85  
DB 631 PARSYAKVPIIGAKKUALN--EVKNVHTV 660

RESULT 3  
A86510  
leucyl tRNA synthetase [imported] - Chlamydomophila pneumoniae (strain J138)  
C;Species: Chlamydomophila pneumoniae, Chlamydia pneumoniae  
C;Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 23-Mar-2001  
C;Accession: A86510  
R;Shirai, M.; Hirakawa, H.; Kimoto, M.; Tabuchi, M.; Kishi, F.; Ouchi, K.; Shiba, T.; In  
Nucleic Acids Res. 28, 2311-2314, 2000  
A;Title: Comparison of whole genome sequences of Chlamydia pneumoniae J138.  
A;Reference number: A86491; MUID:20330349; PMID:10871362  
A;Accession: A86510  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-820 <STO>  
A;Cross-references: GB:BA000008; NID:g8978526; PIDN:BA098363.1; GSPDB:GN00142  
A;Experimental source: strain J138  
C;Genetics:  
A;Gene: leuS  
C;Superfamily: leucine-tRNA ligase

Query Match 9.4%; Score 86.5; DB 2; Length 820;  
Best Local Similarity 24.1%; Pred. No. 5.7;  
Matches 38; Conservative 28; Mismatches 49; Indels 43; Gaps 9;

QY 20 TTSQALPAFTTEIQE-----ASEGPADEVQVFAPANALPARSEAAA-VQPVIGIS 69  
DB 235 TQGSLEAFTTRLDTLGVSLVIAPEHFDLSIV-----SEQRDEVTAVQESLRKS 288

QY 70 QVRVMS-KEKDLGTLGY----VLGITVMV-----IIAIGAGILGYSYKRGKDLKEQ 119  
DB 289 ERDRISSVKTKGVFTGNVAKHPITGNLLPVWISDVVLGYGTGVVMGV-----PA 339

QY 120 HDQKVCEREMQRTLPISAFNPTCEIVDEKTVVHTS 157  
DB 340 HDER--DREPAEM-----FSLPIHEVIDDNGVCIIHSN 369

RESULT 4  
C72113  
leucine-tRNA ligase (EC 6.1.1.4) [similarity] - Chlamydomophila pneumoniae (strains CWL029  
C;Species: Chlamydomophila pneumoniae, Chlamydia pneumoniae  
C;Date: 23-Apr-1999 #sequence\_revision 23-Apr-1999 #text\_change 03-Jun-2002  
C;Accession: C72113; F81557  
R;Kalmann, S.; Mitchell, W.; Marathe, R.; Lammel, C.; Fan, J.; Olinger, L.; Grimwood, J.;  
Nature Genet. 21, 385-389, 1999  
A;Title: Comparative Genomes of Chlamydia pneumoniae and C. trachomatis.  
A;Reference number: A72000; MUID:99206606; PMID:10192388  
A;Accession: C72113  
A;Molecule type: DNA  
A;Residues: 1-820 <ARN>  
A;Cross-references: GB:AE001602; GB:AE001363; NID:g4376416; PIDN:AAD19306.1; PID:g437642  
A;Experimental source: strain CWL029  
R;Read, T.D.; Brunham, R.C.; Shen, C.; Gill, S.R.; Heidelberg, J.F.; White, O.; Hickey,  
C.; Dodson, R.; Gwinn, M.; Nelson, W.; DeBoy, R.; Kolonay, J.; McClarty, G.; Salzberg,  
Nucleic Acids Res. 28, 1397-1406, 2000  
A;Title: Genome sequences of Chlamydia trachomatis MoPn and Chlamydia pneumoniae AR39.

A;Reference number: A81500; MUID:20150255; PMID:10684935  
A;Accession: F81557  
A;Molecule type: DNA  
A;Residues: 1-820 <REA>  
A;Cross-references: GB:AE002219; GB:AE002161; NID:g7189524; PIDN:AAF38433.1; PID:g71895  
A;Experimental source: strain AR39, HL cells  
C;Genetics:  
A;Gene: leuS; CP0618  
C;Superfamily: leucine-tRNA ligase  
C;Keywords: aminoacyl-tRNA synthetase; ligase; protein biosynthesis

Query Match 9.4%; Score 86.5; DB 2; Length 820;  
Best Local Similarity 24.1%; Pred. No. 5.7;  
Matches 38; Conservative 28; Mismatches 49; Indels 43; Gaps 9;

QY 20 TTSQALPAFTTEIQE-----ASEGPADEVQVFAPANALPARSEAAA-VQPVIGIS 69  
DB 235 TQGSLEAFTTRLDTLGVSLVIAPEHFDLSIV-----SEQRDEVTAVQESLRKS 288

QY 70 QVRVMS-KEKDLGTLGY----VLGITVMV-----IIAIGAGILGYSYKRGKDLKEQ 119  
DB 289 ERDRISSVKTKGVFTGNVAKHPITGNLLPVWISDVVLGYGTGVVMGV-----PA 339

QY 120 HDQKVCEREMQRTLPISAFNPTCEIVDEKTVVHTS 157  
DB 340 HDER--DREPAEM-----FSLPIHEVIDDNGVCIIHSN 369

RESULT 5  
T31432  
K-Cl cotransport protein 2, furosemide-sensitive - rat  
C;Species: Rattus norvegicus (Norway rat)  
C;Date: 29-Oct-1999 #sequence\_revision 29-Oct-1999 #text\_change 07-Dec-1999  
C;Accession: T31432  
R;Payne, J.A.; Stevenson, T.J.; Donaldson, I.F.  
J. Biol. Chem. 271, 16245-16252, 1996  
A;Title: Molecular characterization of a putative K-Cl cotransporter in rat brain: a ne  
A;Reference number: Z21031; MUID:96279171; PMID:8663311  
A;Accession: T31432  
A;Status: preliminary; translated from GB/EMBL/DDBJ  
A;Molecule type: mRNA  
A;Residues: 1-1116 <PAP>  
A;Cross-references: EMBL:U55816; NID:g1403708; PID:g1403709; PIDN:AAC52635.1  
A;Experimental source: strain Sprague Dawley; clone ERB10; 5ERB12; brain  
C;Genetics:  
A;Gene: KCC2  
C;Keywords: transmembrane protein

Query Match 9.3%; Score 85.5; DB 2; Length 1116;  
Best Local Similarity 27.1%; Pred. No. 10;  
Matches 32; Conservative 15; Mismatches 44; Indels 27; Gaps 4;

QY 79 KDLGTLGYVLGITMVIITAGIILGYSY-----KRGKDLKEQH-DQKVCEREMQ 130  
DB 872 KDLTFLYHLRTAEVVEVVMHESDISATYKTLVMEQSRQILKQMLTKNEREREIQ 931

QY 131 RIT-----LPLSAFTNPTC---EIVDEKTVVHTTSGTPVDPQEGSTP 169  
DB 932 SITDESRSIRKKNPANTRLNVPETACDNEKEPEEVQLIHDSAPSCSPSSPSP 989

RESULT 6  
B82525  
primosomal protein N' XF2689 [imported] - Xylella fastidiosa (strain 9a5c)  
C;Species: Xylella fastidiosa  
C;Date: 18-Aug-2000 #sequence\_revision 20-Aug-2000 #text\_change 20-Aug-2000  
C;Accession: B82525  
R;Anonymous, The Xylella fastidiosa Consortium of the Organization for Nucleotide Sequ  
Nature 406, 151-157, 2000  
A;Title: The genome sequence of the plant pathogen Xylella fastidiosa.  
A;Reference number: A82515; MUID:20365717; PMID:10910347  
A;Note: for a complete list of authors see reference number A59328 below  
A;Accession: B82525

[illegible]

QY 31 EIQEASEGPGADDEVQVFAPANALPARSEAAVQPVIGISQVRVMSKEKKDLGTL--GYV 88  
 Db 227 EVTPKETPKAPKTETKAKADTEENK-----PSIGVEQTVRVDRDLHMLNIGELV 282  
 QY 89 LGITMNVIIAIGAGIIGYKRGKDLKEHQDQKVCEREMQRTLPLSAFTNFTCSIVD 148  
 Db 283 LKKNLIRI-----YS-----DVEERYDGEKFLLELNQVSSISAVT-----TD 321  
 QY 149 EKTVVVHTSQTVPDQEGSTPLM 171  
 Db 322 LQLAVMKTMRQPVGVKNKFFPM 344

RESULT 10  
 A32657  
 glutamate-cysteine ligase gshl [imported] - Agrobacterium tumefaciens (strain C58, Dupont)  
 C:Species: Agrobacterium tumefaciens  
 C>Date: 11-Jan-2002 #sequence\_revision 11-Jan-2002 #text\_change 18-Nov-2002  
 C:Accession: AB2657  
 R:Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo, I.; Karp, P.; Romero, P.; Zhang, S.  
 Science 294, 2317-2323, 2001  
 A:Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kamm, S.E.W.  
 A:Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.  
 A:Reference number: AB2577; MUID:21608550; PMID:11743193  
 A:Accession: AB2657  
 A>Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-457 <KUR>  
 A:Cross-references: GB:AE008688; PIDN:AA141672.1; PID:g17739016; GSPDB:GN00186  
 A:Experimental source: strain C58 (Dupont)  
 C:Genetics:  
 A:Gene: gshl  
 A:Map position: circular chromosome

Query Match 8.6%; Score 78.5; DB 2; Length 457;  
 Best Local Similarity 32.7%; Pred. No. 16;  
 Matches 33; Conservative 13; Mismatches 26; Indels 29; Gaps 6;

QY 2 GEAGVEKPCEDLRCPETTSQALPAFTTEI---QEASEGPGA-----DEVQVFAPAN 51  
 Db 325 GADGGPWRIC-----ALPAFWGLLYNQALDAADALTADWSFDE--VIALRN 371

QY 52 ALPARSEAAV--QPVIGISQVRVMSKEKKDLGTLG 86  
 Db 372 AVPAKGLAAEIAACKPLLGARQVLDISRTGLKNRKLNGEG 412

RESULT 11  
 G97438  
 glutamate-cysteine ligase precursor, chloroplast (gamma-glutamylcysteine synthetase) (g  
 C:Species: Agrobacterium tumefaciens  
 C>Date: 30-Sep-2001 #sequence\_revision 30-Sep-2001 #text\_change 18-Nov-2002  
 C:Accession: G97438  
 R:Goodner, B.; Hinkle, G.; Gattung, S.; Miller, N.; Blanchard, M.; Quorollo, B.; Goldman, A.; Liu, F.; Wollam, C.; Allinger, M.; Doughty, D.; Scott, C.; Lappas, C.; Markelz, B.;  
 Science 294, 2323-2328, 2001  
 A:Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent Agrobacterium tum  
 A:Reference number: A97359; MUID:21608551; PMID:11743194  
 A:Accession: G97438  
 A>Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-457 <KUR>  
 A:Cross-references: GB:AE007869; PIDN:AAK86464.1; PID:g15155610; GSPDB:GN00169  
 C:Genetics:  
 A:Gene: AGR C 1167  
 A:Map position: circular chromosome

Query Match 8.6%; Score 78.5; DB 2; Length 457;  
 Best Local Similarity 32.7%; Pred. No. 16;  
 Matches 33; Conservative 13; Mismatches 26; Indels 29; Gaps 6;

QY 2 GEAGVEKPCEDLRCPETTSQALPAFTTEI---QEASEGPGA-----DEVQVFAPAN 51  
 Db 325 GADGGPWRIC-----ALPAFWGLLYNQALDAADALTADWSFDE--VIALRN 371

QY 52 ALPARSEAAV--QPVIGISQVRVMSKEKKDLGTLG 86  
 Db 372 AVPAKGLAAEIAACKPLLGARQVLDISRTGLKNRKLNGEG 412

RESULT 12  
 A32555  
 major merozoite surface antigen precursor - Plasmodium chabaudi adami (fragment)  
 C:Species: Plasmodium chabaudi adami  
 C>Date: 07-Sep-1990 #sequence\_revision 07-Sep-1990 #text\_change 11-Jan-2000  
 C:Accession: A32555  
 R:Lew, A.M.; Langford, C.J.; Anders, R.F.; Kemp, D.J.; Saul, A.; Fardoulis, C.; Geysen, P.; Natl. Acad. Sci. U.S.A. 86, 3768-3772, 1989  
 A:Title: A protective monoclonal antibody recognizes a linear epitope in the precursor  
 A:Reference number: A32555; MUID:89264504; PMID:24711191  
 A:Accession: A32555  
 A:Molecule type: mRNA  
 A:Residues: 1-478 <LEW>  
 A:Cross-references: GB:J04568; NID:g160419; PID:g552206  
 C:Superfamily: major merozoite surface antigen  
 C:Keywords: Glycoprotein; surface antigen  
 F:139,299/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 8.6%; Score 78.5; DB 2; Length 478;  
 Best Local Similarity 28.7%; Pred. No. 17;  
 Matches 29; Conservative 11; Mismatches 38; Indels 23; Gaps 3;

QY 3 EAGVPEKPCEDLRCPETTSQALP-----AFTTEIQEASEGPGADEV-----44  
 Db 266 EATQESAPAE-ATTETTPAETPETKEGASTNKSETSEGTAPAEAPSTEVPASPPAT 324

QY 45 -----QVFAPANALPARSEAAVQPVIGISQVRVMSKEKKD 81  
 Db 325 PAAPSASSPAPAPAPQVPTSPQVSGESTNVEGSTQVRAE 365

RESULT 13  
 T22759  
 hypothetical protein F55H12.3 - Caenorhabditis elegans  
 C:Species: Caenorhabditis elegans  
 C>Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 18-Aug-2000  
 C:Accession: T22759  
 R:Dobson, R.  
 submitted to the EMBL Data Library, October 1996  
 A:Reference number: 219610  
 A:Accession: T22759  
 A>Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Residues: 1-2824 <WIL>  
 A:Cross-references: EMBL:Z81091; PIDN:CA803143.1; GSPDB:GN00019; CESP:F55H12.3  
 A:Experimental source: clone F55H12  
 C:Genetics:  
 A:Gene: CESP:F55H12.3  
 A:Map position: 1  
 A:Insertions: 87/1; 98/1; 126/2; 201/3; 343/3; 406/1; 576/3; 656/1; 825/3; 869/1; 909/1; 9  
 7/1; 1755/2; 1800/1; 1850/3; 1896/1; 2003/3; 2035/3; 2082/3; 2119/1; 2144/1; 2200/2; 22  
 C:Superfamily: LDL receptor ligand-binding repeat homology  
 F:243-279/Domain: LDL receptor ligand-binding repeat homology <LDL>

Query Match 8.6%; Score 78.5; DB 2; Length 2824;  
 Best Local Similarity 22.5%; Pred. No. 1.3e+02;  
 Matches 38; Conservative 19; Mismatches 69; Indels 43; Gaps 6;

QY 11 PCEDL-----RCPETTSQALPAFTTEIQEASEGPGADEVQVFAPANALPARSEAAA 61  
 Db 2590 PCSDLSENATSIPTVTCOSTCSDAIPTAGCNQLQNGKSS-----MITRNCPL-----E 2638

QY 62 VQPVIGISQVRVMSKEKKDLGTLGYVLGTMVIIAIGAGIILGYSKRGDLKE-QH 120

```
Db 2639 VTFPGNSNAIKI.....VLGVVFGVLLIIIVLVCFRKQIIAIFRKTDTSNQH 2689
QY 121 DQKVCEREMQRIITPLSAFTNPTCEIVDEKTVVHTSQTFVDPQEGSTP 169
Db 2690 -----VALSHWDNATNNEENQNPSTNTYPRIPQAPIP 2725

RESULT 14
AD0760
diol dehydratase medium chain [imported] - Salmonella enterica subsp. enterica serovar Typhimurium
C:Species: Salmonella enterica subsp. enterica serovar Typhi
A:Note: this species has also been called Salmonella typhi
C:Date: 09-Nov-2001 #sequence_revision 09-Nov-2001 #text_change 22-Jun-2003
C:Accession: AD0760
R:Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Churcher,
  T.; Connor, P.; Cronin, A.; Davis, P.; Davies, R.M.; Dowd, L.; White, N.; Farrar,
  S.; Moule, S.; O'Gaora, P.
  Nature 413, 848-852, 2001
A:Authors: Parry, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.;
  A:Title: Complete genome sequence of a multiple drug resistant Salmonella enterica serovar
  A:Reference number: AB0502; MUID:21534947; PMID:11677608
A:Accession: AD0760
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-224 <PAR>
A:Cross-references: GB:AL513382; PIDN:CAD02402.1; PID:g16503272; GSPDB:GN00176
C:Genetics:
A:Gene: pduD
C:Superfamily: propanediol/glycerol dehydratase, medium subunit

Query Match 8.5%; Score 78; DB 2; Length 224;
Best Local Similarity 27.0%; Pred. No. 7.8;
Matches 34; Conservative 15; Mismatches 37; Indels 40; Gaps 6;

QY 18 PETTSQALPAFTTEIQEASEGDEGVQVFPANALPARSEAAVQVIGISQVRMNSK 77
Db 37 POTAAPAGDGFTEVGEARQGTQDEVII-----AVGPAFLAQTVNIVGL 82
QY 78 EKXDLGTGLGYL-----GITMMVI-----IITAI-----GAGILGYSYKRGKDL 116
Db 83 PHKSI--LREVIAGTEEBGIRAVIRCFKSSDVAFVAVEGNRLSGGISIGI---QSKDT 137
QY 117 KEQHDQ 122
Db 138 TVIHOQ 143

RESULT 15
AI2951
hypothetical protein Atu3215 [imported] - Agrobacterium tumefaciens (strain C58, Dupont)
C:Species: Agrobacterium tumefaciens
C:Date: 11-Jan-2002 #sequence_revision 11-Jan-2002 #text_change 18-Nov-2002
C:Accession: AI2951
R:Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo, L.;
  Gillet, W.; Grant, C.; Guenther, D.; Kutyavin, T.; Levy, R.; Li, M.; McClellan,
  Karp, P.; Romero, P.; Zhang, S.
  Science 294, 2317-2323, 2001
A:Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kamm,
  ster, E.W.
A:Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.
A:Reference number: AB2577; MUID:21608550; PMID:11743193
A:Accession: AI2951
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-267 <KUR>
A:Cross-references: GB:AE008699; PIDN:AAL44031.1; PID:g17741592; GSPDB:GN00187
C:Experimental source: strain C58 (Dupont)
C:Genetics:
A:Gene: Atu3215
A:Map position: linear chromosome

Query Match 8.5%; Score 77.5; DB 2; Length 267;
```

```
Best Local Similarity 21.1%; Pred. No. 11;
Matches 30; Conservative 24; Mismatches 43; Indels 45; Gaps 4;

QY 48 APANALPARSEAA-----AVQVIGISQVRMNSKEKDLGTGLGYLGITMMV 95
Db 53 ALAGLLPARAQVSGDISWSRDSGLFTGRMPG-----RDIGTIFQDTGATLNP 100
QY 96 II-----IAIGAGIILGYSYKRGKDLKQ-----HDQKVCEREMQRIITL 134
Db 101 VITIGEQVAGVVRHGLSWRQGRDLARDLLERVLPHPSHLLSAYPHQLSGGQQRVAI 160
QY 135 PLSAFTNPTCEIVDEKTVVHT 156
Db 161 AALAAAPAILIADATSALDT 182
```

Search completed: March 17, 2004, 07:08:01  
Job time : 14.5276 secs

Thu Mar 18 15:28:10 2004

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 17, 2004, 06:55:43 ; Search time 7.22746 Seconds  
(without alignments)

1289.604 Million cell updates/sec

Title: US-10-057-951-2\_COPY\_85\_263

Perfect score: 916

Sequence: 1 SGAGVPEKPCEDLRCPET.....PVDPEGSGTFLMGAGTPGA 179

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_42.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	86.5	9.4	820	1 SYL CHLPN	Q92930 chlamydia p
2	86	9.4	1165	1 2407 HUMAN	Q9C090 homo sapien
3	85.5	9.3	1115	1 S125 MOUSE	Q91V14 mus musculus
4	85.5	9.3	1116	1 S125 RAT	Q63633 rattus norv
5	81	8.8	499	1 GSHR PLAF7	Q15770 plasmodium
6	81	8.8	483	1 GSHR PLAFK	Q94655 plasmodium
7	80.5	8.8	483	1 TLDD BUCBP	Q89A61 buchnera ap
8	80	8.7	1238	1 JAG2 HUMAN	Q9V219 homo sapien
9	78	8.5	820	1 SYL CHLCV	Q822R7 chlamydophi
10	77.5	8.5	1709	1 SN HUMAN	Q9B222 homo sapien
11	76.5	8.4	425	1 TRIG RAT	P07174 rattus norv
12	76.5	8.4	501	1 GYG2 HUMAN	Q15488 homo sapien
13	76.5	8.4	1116	1 S125 HUMAN	Q9H2X9 homo sapien
14	76	8.3	505	1 SPKD SYNY3	P54735 synechocyst
15	75	8.2	595	1 TNR8 HUMAN	P28908 homo sapien
16	74.5	8.1	747	1 YFGF ECOLI	P77172 escherichia
17	73	8.0	260	1 MTXB METEX	P53595 methyllobact
18	73	8.0	536	1 YENI SCHPO	Q13695 schizosacch
19	73	8.0	676	1 ICPO HSBVJ	P29128 bovine herp
20	73	8.0	698	1 EFG DEIRA	Q9RXK5 deinococcus
21	72.5	7.9	400	1 B3AR MOUSE	P25962 mus musculus
22	72.5	7.9	458	1 DESM XENLA	P23239 xenopus lae
23	72	7.9	646	1 MUI8 HUMAN	P43121 homo sapien
24	72	7.9	952	1 IF41 YEAST	P39935 saccharomyc
25	72	7.9	1781	1 AK12 HUMAN	Q02952 homo sapien
26	71.5	7.8	145	1 YD54 METJA	Q58749 methanococc
27	71.5	7.8	755	1 RRE1 HUMAN	Q92766 homo sapien
28	71.5	7.8	817	1 HUNL MUSDO	Q01778 musca domes
29	71	7.8	255	1 TPIS PHOLL	Q7MYB3 photorhabd
30	71	7.8	341	1 DCUP BRUME	Q8YJ11 bruceella me
31	71	7.8	341	1 DCUP BRUSU	Q8FY24 bruceella su
32	70.5	7.7	241	1 MIAL SARMU	Q26539 sarcocystis
33	70.5	7.7	272	1 IBP1 RAT	P21743 rattus norv

RESULT 1

SYL\_CHLPN

ID SYL\_CHLPN STANDARD; PRT; 820 AA.  
AC Q92930; Q9JQ86;  
DT 30-MAY-2000 (Rel. 39, Created)  
DT 30-MAY-2000 (Rel. 39, Last sequence update)  
DT 10-OCT-2003 (Rel. 42, Last annotation update)  
DE Leucyl-tRNA synthetase (EC 6.1.1.4) (Leucine--tRNA ligase) (LeuRS).  
GN LEUS OR CPN0153 OR CP0618 OR CPB0154.  
OS Chlamydia pneumoniae (Chlamydia pneumoniae).  
OC Bacteria; Chlamydiae; Chlamydiales; Chlamydiaceae; Chlamydia.  
OX NCBI\_TaxID=83558;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=CWLO29;  
RX MEDLINE=99206606; PubMed=10192388;  
RA Kalman S., Mitchell W., Marathe R., Lammel C., Fan J., Hyman R.W.,  
RA Olinger L., Grimwood J., Davis R.W., Stephens R.S.;  
RT "Comparative genomes of Chlamydia pneumoniae and C. trachomatis.";  
RL Nat. Genet. 21:385-389 (1999).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=AR39;  
RX MEDLINE=20150255; PubMed=10694935;  
RA Read T.D., Brunham R.C., Shen C., Gill S.R., Heidelberg J.F., Bass S.,  
RA White O., Hickey E.K., Peterson J., Uterback T., Berry K., Dodson R.,  
RA Linher K., Weidman J., Khouri H., Craven B., Bowman C., Dodson R.,  
RA Gwinn M., Nelson M., DeBoy R., Kolonay J., McClarty G., Salzberg S.L.,  
RA Esen J., Fraser C.M.;  
RT "Genome sequences of Chlamydia trachomatis MoPn and Chlamydia  
RT pneumoniae AR39.";  
RL Nucleic Acids Res. 28:1397-1406 (2000).  
RN [3]  
RP SEQUENCE FROM N.A.  
RC STRAIN=J138;  
RX MEDLINE=20330349; PubMed=10871362;  
RA Shirai M., Hirakawa H., Kimoto M., Tabuchi M., Kishi F., Ouchi K.,  
RA Shiba T., Ishii K., Hattori M., Kuhara S., Nakazawa T.;  
RT "Comparison of whole genome sequences of Chlamydia pneumoniae J138  
RT from Japan and CWLO29 from USA.";  
RL Nucleic Acids Res. 28:2311-2314 (2000).  
RN [4]  
RP SEQUENCE FROM N.A.  
RC STRAIN=TW-183;  
RX Geng M.M., Schumacher A., Muehldorfer I., Bensch K.W., Schaefer K.P.,  
RA Schneider S., Pohl T., Essig A., Marre R., Melchers K.;  
RT "The genome sequence of Chlamydia pneumoniae TW183 and comparison with  
RT other Chlamydia strains based on whole genome sequence analysis.";  
Submitted (MAY-2002) to the EMBL/GenBank/DBJ databases.  
CC -!- CATALYTIC ACTIVITY: ATP + L-leucine + tRNA(Leu) = AMP +  
CC diposphate + L-leucyl-tRNA(Leu).  
CC -!- SUBCELLULAR LOCATION: Cytoplasmic.  
CC -!- SIMILARITY: Belongs to class-I aminoacyl-tRNA synthetase family.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

P29746 drosophila  
Q8rhms fusobacteri  
P12616 actinomycet  
P72689 synechocyst  
P34756 saccharomyc  
P36967 dictyosteli  
O85145 rattus norv  
Q9np33 homo sapien  
Q9xt50 ovis aries  
Q9y8k4 methanosarc  
O8668 sarcocystis  
P47876 mus musculu

ALIGNMENTS



the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

EMBL; AE001602; AAD18306.1; -  
 EMBL; AE002219; AAF38433.1; -  
 EMBL; AP002545; BAA98363.1; -  
 EMBL; AE017157; AAP98087.1; -  
 PIR; A86510; A86510.  
 PIR; C72113; C72113.  
 TIGR; CP0618; -  
 HAMAP; MF\_00049; -; 1.  
 InterPro; IPR002302; Leu-trna-syntyla.  
 InterPro; IPR002300; trna-synt 1a.  
 InterPro; IPR001412; trna-synt 1.  
 InterPro; IPR009008; VALRS ILERS\_edit.  
 Pfam; PF00133; trna-synt 1; 1.  
 PRINTS; PRO0985; TRNASYNTHLEU.  
 TIGRFAMs; TIGR00396; leus\_pact; 1.  
 PROSITE; PS00178; AA\_TRNA\_LIGASE\_1; 1.  
 Aminoacyl-tRNA synthetase; Protein biosynthesis; Ligase; ATP-binding;  
 KW Complete proteome.  
 SITE 40 51 "HIGH" REGION.  
 SITE 601 605 "KWSKS" REGION.  
 BINDING 604 604 ATP (BY SIMILARITY).  
 SEQUENCE 820 AA; 93965 MW; 520369FC098F1926 CRC64;

Query Match 9.4%; Score 86.5; DB 1; Length 820;  
 Best Local Similarity 24.1%; Pred. No. 3.8;  
 Matches 38; Conservative 28; Mismatches 49; Indels 43; Gaps 9;  
 20 TTSQALPATTITQ-----ASEGGADEGVQVFPANALPARSEAAA-VQVIGIS 69  
 235 TQSGLEATTRLDTLILGVSFLVIAPEHPLDSIV-----SEQRDEVTAVQESLRKS 288  
 70 QRVNMS-REKDLGLTGY-----VLGITVMV-----IIIAIGAGILGYSYKRGKDLKEQ 119  
 289 ERDRISVKTGTGVTGNYAKHPITGNLLPVMISDVYVLGYGTGVVMGV-----PA 339  
 120 HDQKVEREMORTLPLSAFTNPTCTIVDEKTVVHTS 157  
 340 HDER--DREPAEM-----FSLPIHEVIDDNGVCIHSN 369

RESULT 2  
 Z407\_HUMAN STANDARD; PRT; 1165 AA.  
 AC Q9C0G0;  
 DT 10-OCT-2003 (Rel. 42, Created)  
 DT 10-OCT-2003 (Rel. 42, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE Zinc finger protein 407 (Fragment).  
 GN ZNF407 OR KIAA1703.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Brain;  
 RX MEDLINE=21082932; PubMed=11214970;  
 RA Nagase T., Kikuno R., Hattori A., Kondo Y., Okumura K., Ohara O.;  
 RT "Prediction of the coding sequences of unidentified human genes. XIX.  
 RT The complete sequences of 100 new cDNA clones from brain which code  
 RT for large proteins in vitro."  
 RL DNA Res. 7:347-355(2000).  
 CC -!- FUNCTION: May function as a transcription factor.  
 CC -!- SUBCELLULAR LOCATION: Nuclear (Potential).  
 CC -!- SIMILARITY: Contains 12 C2H2-type zinc fingers.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration

between the Swiss Institute of Bioinformatics and the EMBL Outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

EMBL; AB051490; BAB21794.1; -  
 HSSP; P08047; ISP2. ZNF407.  
 Genew; HGNC:19904; ZNF C2H2.  
 InterPro; IPR007087; Znf\_C2H2.  
 Pfam; PF00096; zf\_C2H2; 11.  
 SMART; SM00355; Znf\_C2H2; 12.  
 PROSITE; PS00028; ZINC\_FINGER\_C2H2\_1; 5.  
 PROSITE; PS0157; ZINC\_FINGER\_C2H2\_2; 7.  
 Transcription regulation; DNA-binding; Zinc-finger; Metal-binding;  
 KW Nuclear protein; Repeat.  
 NON\_TER 1  
 FT ZN\_FING 331 354 C2H2-TYPE 1 (ATYPICAL).  
 FT ZN\_FING 361 385 C2H2-TYPE 2.  
 FT ZN\_FING 403 426 C2H2-TYPE 3.  
 FT ZN\_FING 454 478 C2H2-TYPE 4 (ATYPICAL).  
 FT ZN\_FING 484 506 C2H2-TYPE 5.  
 FT ZN\_FING 512 535 C2H2-TYPE 6.  
 FT ZN\_FING 545 567 C2H2-TYPE 7.  
 FT ZN\_FING 573 597 C2H2-TYPE 8.  
 FT ZN\_FING 603 625 C2H2-TYPE 9.  
 FT ZN\_FING 631 653 C2H2-TYPE 10.  
 FT ZN\_FING 659 684 C2H2-TYPE 11 (ATYPICAL).  
 FT ZN\_FING 690 713 C2H2-TYPE 12 (ATYPICAL).  
 SQ SEQUENCE 1165 AA; 126980 MW; A37B8A9701F5133E CRC64;

Query Match 9.4%; Score 86; DB 1; Length 1165;  
 Best Local Similarity 24.7%; Pred. No. 6.3;  
 Matches 54; Conservative 30; Mismatches 83; Indels 52; Gaps 13;  
 2 GEAGVPEK-RP-CEDLRCPETTSQALPATTITQEAEGGADEGVQVFPANALPARSE- 58  
 921 GRAGLEQGPAGKDVLI-----QLPG--QEVSHVAADPEAPEIQMFPAQESPAAVEV 972  
 59 -AAAVQVIGISGRVNRMSKEKK-----DLGTIGYVL--GITMVI----- 96  
 973 LTQVHPASAAMASQERAAQVAFKQVGVQLQFAVCDTAAAGQLVKGVTVQVVSEGAVM 1032  
 97 IIAIGAGIILGYSYKRGKDLKE---OHDKVCEREMORTLPLSA--FTNPCTEIV--- 147  
 1033 VAGEGAQIIMQEAQGEHNDLVSDGEISQIIVTEELVQAMVQESSGSGFSGTTHYILTEL 1092  
 148 -----DEXTVVVHTSQTPVDPQE-----GSTPLMQAGTP 177  
 1093 FPGVQDEPGLYSHTVLETADSQELLQAGAT-LGTEAGAP 1130

RESULT 3  
 S125\_MOUSE STANDARD; PRT; 1115 AA.  
 ID S125\_MOUSE  
 AC Q91V14; Q9Z0M7;  
 DT 28-FEB-2003 (Rel. 41, Created)  
 DT 28-FEB-2003 (Rel. 41, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Solute carrier family 12 member 5 (Electroneutral potassium-chloride cotransporter 2) (K-Cl cotransporter 2) (Neuronal K-Cl cotransporter) (mKCC2).  
 DE (mKCC2).  
 GN SLC12A5 OR KCC2.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=ILS, and ISS;  
 RX MEDLINE=21363810; PubMed=11471062;  
 RA Ehringer M.A., Thompson J., Conroy O., Xu Y., Yang F., Canniff J.,

RA Beeson M., Gordon L., Bennett B., Johnson T.E., Sikela J.M.;  
 RT "High-throughput sequence identification of gene coding variants  
 within alcohol-related QTLs.";  
 RL Mamm. Genome 12:657-663 (2001).  
 RN [2]  
 RP SEQUENCE OF 1-204 FROM N.A.  
 RX MEDLINE=99177353; PubMed=10077537;  
 RA Haapa S., Suomalainen S., Eerikainen S., Airaksinen M., Paulin L.,  
 RA Savilahti H.;  
 RT "An efficient DNA sequencing strategy based on bacteriophage Mu in  
 RT vitro DNA transposition reaction.";  
 RL Genome Res. 9:308-315 (1999).  
 RN [3]  
 RP SUBCELLULAR LOCATION  
 RX MEDLINE=21289258; PubMed=11395011;  
 RA Huebner C.A., Stein V., Hermans-Borgmeyer I., Meyer T., Ballanyi K.,  
 RA Jentsch T.G.;  
 RT "Disruption of KCC2 reveals an essential role of K-Cl cotransport  
 RT already in early synaptic inhibition.";  
 RL Neuron 30:515-524 (2001).  
 CC -!- FUNCTION: Mediates electroneutral potassium-chloride cotransport  
 CC in mature neurons. Transport occurs under isotonic conditions, but  
 CC is activated 20-fold by cell swelling. Important for Cl(-)  
 CC homeostasis in neurons.  
 CC -!- SUBUNIT: Homomultimer and heteromultimer with other K-Cl  
 CC cotransporters (By similarity).  
 CC -!- SUBCELLULAR LOCATION: Integral membrane protein. Detected on  
 CC dendrites, but not on axons of spinal cord neurons and at GPHN-  
 CC positive inhibitory synapses.  
 CC -!- DEVELOPMENTAL STAGE: Detected in the ventral horns of the spinal  
 CC cord at E12.5, and throughout the spinal cord at birth.  
 CC -!- SIMILARITY: BELONGS TO THE SLC12A FAMILY OF TRANSPORTERS.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC ENBL; AF332064; AAK56093.1; -  
 CC ENBL; AF332063; AAK56092.1; -  
 CC ENBL; AJ011033; CAA09464.1; -  
 CC MGD; MG11962037; SLC12a5.  
 CC GO; GO:0016021; C:integral to membrane; ISS.  
 CC GO; GO:0005886; C:plasma membrane; IDA.  
 CC GO; GO:0015379; F:potassium:chloride symporter activity; ISS.  
 CC GO; GO:0006873; P:cell ion homeostasis; ISS.  
 CC GO; GO:0006821; P:chloride transport; IPI.  
 CC GO; GO:0007268; P:synaptic transmission; IPI.  
 CC GO; GO:0006810; P:transport; ISS.  
 CC InterPro; IPR002293; AA/rel\_permease1.  
 CC InterPro; IPR004842; KCL cotransp.  
 CC InterPro; IPR000076; KCL cotransp.  
 CC InterPro; IPR004841; Permease region.  
 CC Pfam; PF00324; aa\_permeases; I.  
 CC PRINTS; PR01084; KCLTRANSPT.  
 CC TIGRFAWS; TIGR00930; za30; I.  
 CC Transport; Ion transport; Symport; Potassium; Potassium transport;  
 CC Transmembrane.  
 CC DOMAIN 1 110 CYTOPLASMIC (POTENTIAL).  
 CC TRANSMEM 111 131 POTENTIAL.  
 CC TRANSMEM 133 153 POTENTIAL.  
 CC DOMAIN 154 171 CYTOPLASMIC (POTENTIAL).  
 CC TRANSMEM 172 192 POTENTIAL.  
 CC TRANSMEM 194 214 POTENTIAL.  
 CC DOMAIN 215 231 CYTOPLASMIC (POTENTIAL).  
 CC TRANSMEM 232 252 POTENTIAL.  
 CC TRANSMEM 255 275 POTENTIAL.  
 CC DOMAIN 276 395 CYTOPLASMIC (POTENTIAL).  
 CC TRANSMEM 396 416 POTENTIAL.  
 CC TRANSMEM 436 456 POTENTIAL.

FT DOMAIN 457 473 CYTOPLASMIC (POTENTIAL).  
 FT TRANSMEM 474 494 POTENTIAL.  
 FT TRANSMEM 547 567 POTENTIAL.  
 FT DOMAIN 568 607 CYTOPLASMIC (POTENTIAL).  
 FT TRANSMEM 608 628 POTENTIAL.  
 FT TRANSMEM 825 845 POTENTIAL.  
 FT DOMAIN 846 1115 CYTOPLASMIC (POTENTIAL).  
 FT CARBOHYD 419 419 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 810 810 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 SQ SEQUENCE 1115 AA; 123587 MW; DD506AC24D8492B4 CRC64;  
 Query Match 9.3%; Score 85.5; DB 1; Length 1115;  
 Best Local Similarity 27.1%; Pred. No. 6.6;  
 Matches 32; Conservative 15; Mismatches 44; Indels 27; Gaps 4;  
 QY 79 KDLGTGLGVVIGITMVIHIIAIGAGIILGVSY-----KRGKDLKXQH-DQKVCEREMQ 130  
 DB 872 KDLTFLHLRITAEVVEVMEHSDISATYVEKTLVMEQSRQILKQMLTKVEREREIQ 931  
 QY 131 RIT-----LPLSAFTNPTC---EIVDEKTVVHTSQTPVDPQEGSTP 169  
 DB 932 SITDESGSIRRNKPNRPLRLNVPETACDNEEKPEEVQLIHQDSAPSPSSPSP 989  
 RESULT 4  
 S125 RAT STANDARD; PRT; 1116 AA.  
 AC Q63633;  
 DT 28-FEB-2003 (Rel. 41, Created)  
 DT 28-FEB-2003 (Rel. 41, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Solute carrier family 12 member 5 (Electroneutral potassium-chloride  
 DE cotransporter 2) (K-Cl cotransporter 2) (Neuronal K-Cl cotransporter)  
 DE (Furosemide-sensitive K-Cl cotransporter) (rKCC2).  
 GN SLC12A5 OR KCC2.  
 OS Rattus norvegicus (Rat).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
 OX NCBI\_TaxID=10116;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Sprague-Dawley; TISSUE=Brain;  
 RX MEDLINE=96279171; PubMed=8663311;  
 RA Payne J.A., Stevenson T.J., Donaldson L.F.;  
 RT "Molecular characterization of a putative K-Cl cotransporter in rat  
 RL brain. A neuronal-specific isoform.";  
 RL J. Biol. Chem. 271:16245-16252 (1996).  
 RN [2]  
 RP FUNCTION.  
 RX MEDLINE=99127889; PubMed=9930699;  
 RA Rivera C., Voipio J., Payne J.A., Ruusuvaara E., Lahtinen H.,  
 RA Lamsa K., Pirvola U., Saarima M., Kaila K.;  
 RT "The K+/Cl- co-transporter KCC2 renders GABA hyperpolarizing during  
 RT neuronal maturation.";  
 RL Nature 397:251-255 (1999).  
 RN [3]  
 RP SUBUNIT.  
 RX MEDLINE=2151256; PubMed=11551954;  
 RA Casula S., Shmukler B.E., Wilhelm S., Stuart-Tilley A.K., Su W.,  
 RA Chernova M.N., Brugnara C., Alper S.L.;  
 RT "A dominant negative mutant of the KCC1 K-Cl cotransporter: both N-  
 RT and C-terminal cytoplasmic domains are required for K-Cl cotransport  
 RT activity.";  
 RL J. Biol. Chem. 276:41870-41878 (2001).  
 CC -!- FUNCTION: Mediates electroneutral potassium-chloride cotransport  
 CC in mature neurons. Transport occurs under isotonic conditions, but  
 CC is activated 20-fold by cell swelling. Important for Cl(-)  
 CC homeostasis in neurons. Necessary for the ontogenic change in  
 CC response to gamma-aminobutyric acid (GABA) from depolarization to  
 CC hyperpolarization during neuronal development.  
 CC -!- SUBUNIT: Homomultimer and heteromultimer with other K-Cl  
 CC cotransporters (By similarity).  
 CC -!- SUBCELLULAR LOCATION: Integral membrane protein.

CC -!- TISSUE SPECIFICITY: Highly expressed in brain. Not detected in  
CC other tissues. Highly expressed in pyramidal neurons and in  
CC neurons throughout the cortex, hippocampus, the granular layer of  
CC the cerebellum and in groups of neurons throughout the brainstem.  
CC Barely detectable in dorsal-root ganglions.  
CC -!- DEVELOPMENTAL STAGE: Detected in thalamus, but not in hippocampus.  
CC and neocortex at E20. At birth barely detectable in hippocampus.  
CC Expression increases steeply from day 5 to day 9 and then  
CC stabilizes at adult levels.  
CC -!- SIMILARITY: BELONGS TO THE SLC12A FAMILY OF TRANSPORTERS.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; U55816; AAC52635.1; ..  
CC PIR; T31432; T31432.  
CC GO; GO:0016021; C: integral to membrane; ISS.  
CC GO; GO:0015379; F: potassium:chloride symporter activity; ISS.  
CC GO; GO:0006873; P: cell ion homeostasis; ISS.  
CC GO; GO:0006810; P: transport; ISS.  
CC InterPro; IPR002293; AA/rel\_permease1.  
CC InterPro; IPR004842; KCL\_cotranspt.  
CC InterPro; IPR000076; KCL\_cotranspt.  
CC InterPro; IPR004841; Permease region.  
CC Pfam; PF00324; aa\_permeases; I.  
CC PRINTS; PR01081; KCLTRANSPORT.  
CC TIGRPFAMs; TIGR00930; 2a30.1.  
CC Transport; Ion transport; Symport; Potassium; Potassium transport;  
KW Transmembrane.  
FT DOMAIN 1 110 CYTOPLASMIC (POTENTIAL).  
FT TRANSMEM 111 131 POTENTIAL.  
FT TRANSMEM 133 153 POTENTIAL.  
FT DOMAIN 154 171 CYTOPLASMIC (POTENTIAL).  
FT TRANSMEM 172 192 POTENTIAL.  
FT TRANSMEM 194 214 POTENTIAL.  
FT DOMAIN 215 231 CYTOPLASMIC (POTENTIAL).  
FT TRANSMEM 232 252 POTENTIAL.  
FT TRANSMEM 255 275 POTENTIAL.  
FT DOMAIN 276 395 CYTOPLASMIC (POTENTIAL).  
FT TRANSMEM 396 416 POTENTIAL.  
FT TRANSMEM 436 456 POTENTIAL.  
FT DOMAIN 457 473 CYTOPLASMIC (POTENTIAL).  
FT TRANSMEM 474 494 POTENTIAL.  
FT TRANSMEM 547 567 POTENTIAL.  
FT DOMAIN 568 607 CYTOPLASMIC (POTENTIAL).  
FT TRANSMEM 608 628 POTENTIAL.  
FT TRANSMEM 825 845 POTENTIAL.  
FT DOMAIN 846 1116 CYTOPLASMIC (POTENTIAL).  
FT CARBOHYD 419 419 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 810 810 N-LINKED (GLCNAC. . .) (POTENTIAL).  
SQ SEQUENCE 1116 AA; 123563 MW; 8E2FDCDD7FDE2F6 CRC64;  
  
Query Match 9.3%; Score 85.5; DB 1; Length 1116;  
Best Local Similarity 27.1%; Pred. No. 6.6;  
Matches 32; Conservative 15; Mismatches 44; Indels 27; Gaps 4;  
  
QY 79 KIDLGTGLVGLITMVIITIIAGIILGYSGY-----XRGKDLKEQH-DQKVCEREMQ 130  
DB 872 KDLTFLYHLRITAEVVEVMEHSDISAVTYEKLWQEQSQILKQMLTKNEREREIQ 931  
  
QY 131 RIT-----IPLSFTNPTC---HIVDEKVVVHTSQTFVDPQEGSTP 169  
DB 932 SITDESIRKRNKPANTRLANVPEETACDNEEKPEEVQLIHDSAPSCSPSSPSP 989  
  
RESULT 5  
GSHR\_PLAF7  
ID GSHR\_PLAF7 STANDARD; PRT; 499 AA.

AC 015770;  
DT 16-OCT-2001 (Rel. 40, Created)  
DT 16-OCT-2001 (Rel. 40, Last sequence update)  
DT 15-MAR-2004 (Rel. 43, Last annotation update)  
DE Glutathione reductase (EC 1.8.1.7) (GR) (Grase).  
GN GR3.  
OS Plasmodium falciparum (isolate 3D7).  
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.  
OX NCBI\_TaxID=36329;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Gilberger T.-W., Walter R.D., Mueller S.;  
RL Submitted (OCT-1997) to the EMBL/GenBank/DBSJ databases.  
CC -!- FUNCTION: Maintain high levels of reduced glutathione in the  
CC cytosol (By similarity).  
CC -!- CATALYTIC ACTIVITY: 2 glutathione + NADP(+) = glutathione  
CC disulfide + NADPH.  
CC -!- COFACTOR: Binds 1 FAD per subunit (By similarity).  
CC -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).  
CC -!- MISCELLANEOUS: The active site is a redox-active disulfide bond.  
CC -!- SIMILARITY: Belongs to class-I pyridine nucleotide-disulfide  
CC oxidoreductase family.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; AF027825; AB884117.1; ..  
CC HSBF; F00390; IALG.  
CC InterPro; IPR001327; FAD\_pyr\_redox.  
CC InterPro; IPR000815; Hg\_reductase.  
CC InterPro; IPR001100; Pyr\_redox.  
CC InterPro; IPR004099; pyr\_redox\_dim.  
CC InterPro; IPR001013; Pyridine\_redox\_2.  
CC Pfam; PF02852; pyr\_redox; 1.  
CC Pfam; PF02852; pyr\_redox\_dim; 1.  
CC PRINTS; PR00368; FADPNR.  
CC PRINTS; PR00945; HGRDTASE.  
CC PRINTS; PR00411; PNDRDTASEII.  
CC PRINTS; PR00469; PNDRDTASEII.  
CC PRODOM; PD000139; FAD\_pyr\_redox; 1.  
CC PROSITE; PS00076; PYRIDINE\_REDOX\_1; 1.  
KW Redox-active center; Oxidoreductase; Flavoprotein; FAD; NADP.  
FT INIT MET 0 BY SIMILARITY  
FT NP\_BIND 31 39 FAD (ADP PART) (BY SIMILARITY).  
FT DISULFID 39 44 REDOX-ACTIVE (BY SIMILARITY).  
FT ACT SITE 484 BY SIMILARITY.  
SQ SEQUENCE 499 AA; 56288 MW; 43CE0251E7B8244 CRC64;  
  
Query Match 8.8%; Score 81; DB 1; Length 499;  
Best Local Similarity 24.5%; Pred. No. 7;  
Matches 34; Conservative 25; Mismatches 64; Indels 16; Gaps 7;  
  
QY 29 TTEIQEASGEGADEVQVPAPNALPARSEAAVPIGVSQVR---MNSKEKDLGT 84  
DB 123 TKNNKNGKPLNEE--ILEGRNLIAGVGNKVPFPVPGIENTISSDEFFNKEKKIGI 180  
  
QY 85 LGY-VLGITMVIITIIAGIILGYSGYKKGDLKEHQDQ---KVCEREMQRIITLPLSFT 140  
DB 181 VGSYIAVELINIVKRLG---IDSYIFARGNRLRKFDSEVINLVLENDKKNNINIVTFA 237  
  
QY 141 NPTCEI--VDEKVVVHTS 157  
DB 238 D-VVEIKKVSCKNLISHL 255  
  
RESULT 6  
GSHR\_PLAFK  
ID GSHR\_PLAFK STANDARD; PRT; 499 AA.

Q94655; 16-OCT-2001 (Rel. 40, Created)  
DT 16-OCT-2001 (Rel. 40, Last sequence update)  
DT 15-MAR-2004 (Rel. 43, Last annotation update)  
DE Glutathione reductase (EC 1.8.1.7) (GR) (Grase).  
GN GR2.  
OS Plasmodium falciparum (isolate K1 / Thailand).  
OC Eukaryota; Alveolata; Apicomplexa; Haemosporidia; Plasmodium.  
OX NCBI\_TaxID=5839;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=96370813; PubMed=8774709;  
RA Faerber P.M., Becker K., Mueller S.;  
RT "Molecular cloning and characterization of a putative glutathione  
RT reductase gene, the pfgr2 gene, from Plasmodium falciparum.";  
RL Eur. J. Biochem. 239:655-661(1996).  
RN [2]  
RP SEQUENCE OF 1-15, AND CHARACTERIZATION.  
RX MEDLINE=96202957; PubMed=8631352;  
RA Krauth-Siegel R.L., Muller J.C., Lottspeich F., Schirmer R.H.;  
RT "Glutathione reductase and glutamate dehydrogenase of Plasmodium  
RT falciparum, the causative agent of tropical malaria.";  
RL Eur. J. Biochem. 235:345-350(1996).  
CC -!- FUNCTION: Maintains high levels of reduced glutathione in the  
CC cytosol (By similarity).  
CC -!- CATALYTIC ACTIVITY: 2 glutathione + NADP(+) = glutathione  
CC disulfide + NADPH.  
CC -!- COFACTOR: Binds 1 FAD per subunit (By similarity).  
CC -!- SUBUNIT: Homodimer.  
CC -!- SUBCELLULAR LOCATION: Cytoplasmic.  
CC -!- MISCELLANEOUS: The active site is a redox-active disulfide bond.  
CC -!- SIMILARITY: Belongs to class-I pyridine nucleotide-disulfide  
CC oxidoreductase family.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/  
CC or send an email to license@isb-sib.ch).  
CC  
CC EMBL; X93462; CAA63747.1; -.  
DR HSP; P00390; IALG.  
DR InterPro: IPR001327; FAD pyr redox.  
DR InterPro: IPR000815; Hg\_reductase.  
DR InterPro: IPR001100; Pyr\_redox.  
DR InterPro: IPR004099; Pyr\_redox\_dim.  
DR InterPro: IPR00103; Pyridine\_redox\_2.  
DR Pfam: PF00070; Pyr\_redox; 1.  
DR Pfam: PF02852; Pyr\_redox\_dim; 1.  
DR PRINTS; P00368; FADNR.  
DR PRINTS; P00945; HGRDTASE.  
DR PRINTS; P00411; PNDRTASEI.  
DR PRINTS; P00469; PNDRTASEII.  
DR PRODOM: PD000139; FAD pyr redox; 1.  
DR PROSITE; PS00076; PYRIDINE REDOX 1; 1.  
DR Redox-active center; Oxidoreductase; Flavoprotein; FAD; NADP.  
FT INIT MET 0 0  
FT NP BIND 31 39 FAD (ADP PART) (BY SIMILARITY).  
FT DISULFID 39 44 REDOX-ACTIVE (BY SIMILARITY).  
FT ACT\_SITE 484 484 BY SIMILARITY.  
FT CDS 484 484  
SQ SEQUENCE 499 AA; 56430 MW; 6B229901ECA095F7 CRC64;  
Query Match 8.8%; Score 81; DB 1; Length 499;  
Best Local Similarity 24.5%; Pred. No. 7;  
Matches 34; Conservative 25; Mismatches 64; Indels 16; Gaps 7;  
QY 29 TTEIQASGEGGADEVQVFPANALPARSEAAVQVIGISQVR-----MNSKEKOLGT 84  
Db 123 TKDNKKNDGPNLEE--ILEGRNIIAVGNKVPFPVFKGIENTISSDEFNFNKESKIGI 180  
QY 85 LGY-VLGIITWVIIIAIGAGIILGYSYKRGKDLKEHQDQ---KVCREMQRITLPLSFT 140  
Q94655; 16-OCT-2001 (Rel. 40, Created)  
DT 16-OCT-2001 (Rel. 40, Last sequence update)  
DT 15-MAR-2004 (Rel. 43, Last annotation update)  
DE Glutathione reductase (EC 1.8.1.7) (GR) (Grase).  
GN GR2.  
OS Plasmodium falciparum (isolate K1 / Thailand).  
OC Eukaryota; Alveolata; Apicomplexa; Haemosporidia; Plasmodium.  
OX NCBI\_TaxID=5839;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=96370813; PubMed=8774709;  
RA Faerber P.M., Becker K., Mueller S.;  
RT "Molecular cloning and characterization of a putative glutathione  
RT reductase gene, the pfgr2 gene, from Plasmodium falciparum.";  
RL Eur. J. Biochem. 239:655-661(1996).  
RN [2]  
RP SEQUENCE OF 1-15, AND CHARACTERIZATION.  
RX MEDLINE=96202957; PubMed=8631352;  
RA Krauth-Siegel R.L., Muller J.C., Lottspeich F., Schirmer R.H.;  
RT "Glutathione reductase and glutamate dehydrogenase of Plasmodium  
RT falciparum, the causative agent of tropical malaria.";  
RL Eur. J. Biochem. 235:345-350(1996).  
CC -!- FUNCTION: Maintains high levels of reduced glutathione in the  
CC cytosol (By similarity).  
CC -!- CATALYTIC ACTIVITY: 2 glutathione + NADP(+) = glutathione  
CC disulfide + NADPH.  
CC -!- COFACTOR: Binds 1 FAD per subunit (By similarity).  
CC -!- SUBUNIT: Homodimer.  
CC -!- SUBCELLULAR LOCATION: Cytoplasmic.  
CC -!- MISCELLANEOUS: The active site is a redox-active disulfide bond.  
CC -!- SIMILARITY: Belongs to class-I pyridine nucleotide-disulfide  
CC oxidoreductase family.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/  
CC or send an email to license@isb-sib.ch).  
CC  
CC EMBL; X93462; CAA63747.1; -.  
DR HSP; P00390; IALG.  
DR InterPro: IPR001327; FAD pyr redox.  
DR InterPro: IPR000815; Hg\_reductase.  
DR InterPro: IPR001100; Pyr\_redox.  
DR InterPro: IPR004099; Pyr\_redox\_dim.  
DR InterPro: IPR00103; Pyridine\_redox\_2.  
DR Pfam: PF00070; Pyr\_redox; 1.  
DR Pfam: PF02852; Pyr\_redox\_dim; 1.  
DR PRINTS; P00368; FADNR.  
DR PRINTS; P00945; HGRDTASE.  
DR PRINTS; P00411; PNDRTASEI.  
DR PRINTS; P00469; PNDRTASEII.  
DR PRODOM: PD000139; FAD pyr redox; 1.  
DR PROSITE; PS00076; PYRIDINE REDOX 1; 1.  
DR Redox-active center; Oxidoreductase; Flavoprotein; FAD; NADP.  
FT INIT MET 0 0  
FT NP BIND 31 39 FAD (ADP PART) (BY SIMILARITY).  
FT DISULFID 39 44 REDOX-ACTIVE (BY SIMILARITY).  
FT ACT\_SITE 484 484 BY SIMILARITY.  
FT CDS 484 484  
SQ SEQUENCE 499 AA; 56430 MW; 6B229901ECA095F7 CRC64;  
Query Match 8.8%; Score 81; DB 1; Length 499;  
Best Local Similarity 24.5%; Pred. No. 7;  
Matches 34; Conservative 25; Mismatches 64; Indels 16; Gaps 7;  
QY 29 TTEIQASGEGGADEVQVFPANALPARSEAAVQVIGISQVR-----MNSKEKOLGT 84  
Db 123 TKDNKKNDGPNLEE--ILEGRNIIAVGNKVPFPVFKGIENTISSDEFNFNKESKIGI 180  
QY 85 LGY-VLGIITWVIIIAIGAGIILGYSYKRGKDLKEHQDQ---KVCREMQRITLPLSFT 140  
Q94655; 16-OCT-2001 (Rel. 40, Created)  
DT 16-OCT-2001 (Rel. 40, Last sequence update)  
DT 15-MAR-2004 (Rel. 43, Last annotation update)  
DE Glutathione reductase (EC 1.8.1.7) (GR) (Grase).  
GN GR2.  
OS Plasmodium falciparum (isolate K1 / Thailand).  
OC Eukaryota; Alveolata; Apicomplexa; Haemosporidia; Plasmodium.  
OX NCBI\_TaxID=5839;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=96370813; PubMed=8774709;  
RA Faerber P.M., Becker K., Mueller S.;  
RT "Molecular cloning and characterization of a putative glutathione  
RT reductase gene, the pfgr2 gene, from Plasmodium falciparum.";  
RL Eur. J. Biochem. 239:655-661(1996).  
RN [2]  
RP SEQUENCE OF 1-15, AND CHARACTERIZATION.  
RX MEDLINE=96202957; PubMed=8631352;  
RA Krauth-Siegel R.L., Muller J.C., Lottspeich F., Schirmer R.H.;  
RT "Glutathione reductase and glutamate dehydrogenase of Plasmodium  
RT falciparum, the causative agent of tropical malaria.";  
RL Eur. J. Biochem. 235:345-350(1996).  
CC -!- FUNCTION: Maintains high levels of reduced glutathione in the  
CC cytosol (By similarity).  
CC -!- CATALYTIC ACTIVITY: 2 glutathione + NADP(+) = glutathione  
CC disulfide + NADPH.  
CC -!- COFACTOR: Binds 1 FAD per subunit (By similarity).  
CC -!- SUBUNIT: Homodimer.  
CC -!- SUBCELLULAR LOCATION: Cytoplasmic.  
CC -!- MISCELLANEOUS: The active site is a redox-active disulfide bond.  
CC -!- SIMILARITY: Belongs to class-I pyridine nucleotide-disulfide  
CC oxidoreductase family.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/  
CC or send an email to license@isb-sib.ch).  
CC  
CC EMBL; X93462; CAA63747.1; -.  
DR HSP; P00390; IALG.  
DR InterPro: IPR001327; FAD pyr redox.  
DR InterPro: IPR000815; Hg\_reductase.  
DR InterPro: IPR001100; Pyr\_redox.  
DR InterPro: IPR004099; Pyr\_redox\_dim.  
DR InterPro: IPR00103; Pyridine\_redox\_2.  
DR Pfam: PF00070; Pyr\_redox; 1.  
DR Pfam: PF02852; Pyr\_redox\_dim; 1.  
DR PRINTS; P00368; FADNR.  
DR PRINTS; P00945; HGRDTASE.  
DR PRINTS; P00411; PNDRTASEI.  
DR PRINTS; P00469; PNDRTASEII.  
DR PRODOM: PD000139; FAD pyr redox; 1.  
DR PROSITE; PS00076; PYRIDINE REDOX 1; 1.  
DR Redox-active center; Oxidoreductase; Flavoprotein; FAD; NADP.  
FT INIT MET 0 0  
FT NP BIND 31 39 FAD (ADP PART) (BY SIMILARITY).  
FT DISULFID 39 44 REDOX-ACTIVE (BY SIMILARITY).  
FT ACT\_SITE 484 484 BY SIMILARITY.  
FT CDS 484 484  
SQ SEQUENCE 499 AA; 56430 MW; 6B229901ECA095F7 CRC64;  
Query Match 8.8%; Score 80.5; DB 1; Length 483;  
Best Local Similarity 20.9%; Pred. No. 7.6;  
Matches 51; Conservative 18; Mismatches 60; Indels 115; Gaps 10;  
QY 17 CPETTSQALPAFT-----TEIQ-----EASEGGADEVQV 46  
Db 110 CPLTTTTRTKPIYTSNPLTSIDTQKLELRNCVAKYDHRVSVNALSSESYDEVLI 169  
QY 47 FAPANALPARSEAAVQVIGISQVRMNSKEKOLGTGLGYVLGITWVIIIAIGAGIIL 106  
Db 170 ATSDGNL-----AADIRPLRLSINVLVNDKGFERG-----VSGGSGRS 209  
QY 107 GYSY-----KRGDLKEHQKQVCE-----REMQRITLPL-----136  
Db 210 GYSFPLNKHKSGKILAEFY---ACEAARIALINLSAQEAPSGTFPVVLGSGWGVLLHEA 266  
QY 137 -----SAFTN-----PTCEIVDEKTVVVVHTSQTVPDPOEGSTPLMGQA 174  
Db 267 VGHLEGDFNRQGTSTVFANKIGRQVASELCTIVDDGTGLKDRGSLTIDDE-----316  
QY 175 GTPG 178  
Db 317 GTPG 320  
RESULT 8  
JAG2\_HUMAN  
ID JAG2\_HUMAN STANDARD; PRT; 1238 AA.

EMBL	AF111170	AA15562.1	-
EMBL	Y14330	CAA74705.1	-
DR	HSSP	P00743	ICCF
DR	GenBank	HGNC:6189	JAG2
DR	MM	602570	-
DR	GO	GO:0005887	C:integral to plasma membrane; ISS.
DR	GO	GO:0008083	F:growth factor activity; IDA.
DR	GO	GO:0005112	F:Notch binding; IPI.
DR	GO	GO:0007049	P:cell cycle; NAS.
DR	GO	GO:0030154	P:cell differentiation; IDA.
DR	GO	GO:0001709	P:cell fate determination; NAS.
DR	GO	GO:0007267	P:cell-cell signaling; ISS.
DR	GO	GO:0009912	P:hair cell fate commitment; ISS.
DR	GO	GO:0007605	P:hearing; ISS.
DR	GO	GO:0030326	P:limb morphogenesis; ISS.
DR	GO	GO:0007219	P:N signaling pathway; NAS.
DR	GO	GO:0030334	P:regulation of cell migration; NAS.
DR	GO	GO:0042127	P:regulation of cell proliferation; IDA.
DR	GO	GO:0007293	P:spermatogenesis; IEP.
DR	GO	GO:0030217	P:cell differentiation; IDA.
DR	GO	GO:0045061	P:thymic T-cell selection; IDA.
DR	InterPro	IPR000152	Aex_hydroxyl_S.
DR	InterPro	IPR001774	DSL.
DR	InterPro	IPR000742	EGF 2.
DR	InterPro	IPR001881	EGF Ca.
DR	InterPro	IPR001438	EGF II.
DR	InterPro	IPR006209	EGF-like.
DR	InterPro	IPR001007	VWF_C.
DR	Pfam	PF04414	DSL; 1.
DR	Pfam	PF00008	EGF; 14.
DR	PRINTS	PR00010	EGFBLOOD.
DR	SMART	SM00051	DSL; 1.
DR	SMART	SM00179	EGF CA; 9.
DR	SMART	SM00214	VWC; 1.
DR	PROSITE	PS00010	ASX_HYDROXYL; 10.
DR	PROSITE	PS00022	EGF_1; 16.
DR	PROSITE	PS01186	EGF_2; 12.
DR	PROSITE	PS00026	EGF_3; 15.
DR	PROSITE	PS01187	EGF CA; 7.
DR	PROSITE	PS0184	VWF_2; 1.
DR	Calcium-binding	EGF-like domain	Glycoprotein; Developmental protein; Repeat; Signal; Transmembrane; Alternative splicing.
PT	SIGNAL	1	POTENTIAL.
PT	CHAIN	27	JAGGED 2.
PT	DOMAIN	27	EXTRACELLULAR (POTENTIAL).
PT	TRANSMEM	1081	POTENTIAL.
PT	DOMAIN	1106	CYTOLASMIC (POTENTIAL).
PT	DOMAIN	178	DSL.
PT	DOMAIN	241	EGF-LIKE 1.
PT	DOMAIN	275	EGF-LIKE 2.
PT	DOMAIN	307	EGF-LIKE 3.
PT	DOMAIN	347	EGF-LIKE 4.
PT	DOMAIN	385	EGF-LIKE 5.
PT	DOMAIN	423	EGF-LIKE 6.
PT	DOMAIN	461	EGF-LIKE 7.
PT	DOMAIN	498	EGF-LIKE 8.
PT	DOMAIN	536	EGF-LIKE 9.
PT	DOMAIN	574	EGF-LIKE 10 (ATYPICAL).
PT	DOMAIN	636	EGF-LIKE 11, CALCIUM-BINDING (POTENTIAL).
PT	DOMAIN	674	EGF-LIKE 12, CALCIUM-BINDING (POTENTIAL).
PT	DOMAIN	712	EGF-LIKE 13.
PT	DOMAIN	751	EGF-LIKE 14.
PT	DOMAIN	789	EGF-LIKE 15.
PT	DOMAIN	827	EGF-LIKE 16, CALCIUM-BINDING (POTENTIAL).
PT	DOMAIN	870	VWF.
PT	DISULFID	245	BY SIMILARITY.
PT	DISULFID	249	BY SIMILARITY.
PT	DISULFID	264	BY SIMILARITY.
PT	DISULFID	276	BY SIMILARITY.
PT	DISULFID	282	BY SIMILARITY.
PT	DISULFID	295	BY SIMILARITY.
PT	DISULFID	311	BY SIMILARITY.
PT	DISULFID	317	BY SIMILARITY.

Query Match  
Best Local Similarity 8.7%; Score 80; DB 1; Length 1238;  
Matches 55; Conservative 20; Mismatches 77; Indels 94; Gaps 11;

QY 2 GEAGV-PEKPC-----EDLRC-----PETS-----GALPAFTTEIQ 33  
DB 947 GCGAEEPPSTCLPRSHLDNNCARLTHTNRDHPGCTTGAICSGISLPATRAVAR 1006

QY 34 E-----ASEGFGADEVOV-PAPANALPARS--EAAAVQPVIGISQRVRMNSKEKD 81  
DB 1007 DRLLVLCDRASSGASAVEAVSFSPARDLPDSSLIOGAHAIVAAAITOR----- 1056

QY 82 LGTLGVGTITMWIIIAIGAILG-----YSYXRGKDLKE 118  
DB 1057 -GNSSULLAVETKVETVTGGSSTGLLVFLVCGFSLWILACVVLCVWTGRERK---- 1111

QY 119 QHDQKCVEREMOIRIT-----LPLSAFTNTPTCEIVDKTKTVVHVSQTTPVDFQEGSTP 169  
DB 1112 -----FRERSLPREESANNQWAPLNPIRNPIERPQGHDKVLYQCQNFTPTPPERRADEA 1164

QY 170 LMGOAG 175  
DB 1165 LPPGAG 1170

RESULT 9  
SYL\_CHLCV STANDARD; PRT; 820 AA.

ID SYL CHLCV  
AC Q822R7;  
DT 10-OCT-2003 (Rel. 42, Created)  
DT 10-OCT-2003 (Rel. 42, Last sequence update)  
DT 10-OCT-2003 (Rel. 42, Last annotation update)  
DE Leucyl-tRNA synthetase (EC 6.1.1.4) (Leucine--tRNA ligase) (Leurs).  
LEUS OR CCA00612.  
OS Chlamydomophila caviae.  
OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydiales; Chlamydia phila.  
OX NCBI\_TaxID=83557;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=GPIC;  
RX MEDLINE=22569155; PubMed=12682364;  
RA Read T.D., Myers G.S.A., Brunham R.C., Nelson W.C., Paulsen I.T.,  
RA Heidelberg J., Holtzapple E., Khouri H., Federova N.B., Carty H.A.,  
RA Umayam L.A., Haft D.H., Peterson J., Beanan M.J., White O.,  
RA Salzberg S.L., Hsia R.-C., McClarty G., Rank R.G., Bavovill P.M.,  
RA Fraser C.M.;  
RT "Genome sequence of Chlamydomophila caviae (Chlamydia psittaci GPIC):  
examining the role of niche-specific genes in the evolution of the  
Chlamydiaceae.";  
RL Nucleic Acids Res. 31:2134-2147(2003).  
CC -!- CATALYTIC ACTIVITY: ATP + L-leucine + tRNA(Leu) = AMP +  
diphosphate + L-leucyl-tRNA(Leu).  
CC -!- SUBCELLULAR LOCATION: Cytoplasmic.  
CC -!- SIMILARITY: Belongs to class-I aminoacyl-tRNA synthetase family.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (see http://www.isb-sib.ch/announce/  
or send an email to license@isb-sib.ch).

DR EMBL; AE016996; AAP05354.1; --  
DR TIGR; CCA00612; --  
DR HAMAP; MF\_00049; -- 1.  
DR InterPro; IPR002302; Leu-tRNA synthetase.  
DR InterPro; IPR002300; tRNA-synt\_1a.  
DR InterPro; IPR001412; tRNA-synt\_i.  
DR InterPro; IPR009008; ValRS\_IleRS\_edit.  
DR Pfam; PF00133; tRNA-synt\_1; 1.  
DR PRINTS; PR00985; TRNASYNTHLEU.  
DR TIGRFAMs; TIGR00396; leus bact; 1.  
DR PROSITE; PS00178; AA TRNA\_LIGASE I; 1.  
KW Aminoacyl-tRNA synthetase; Protein biosynthesis; Ligase; ATP-binding;  
Complete proteome.  
FT SITE 40 51 "HIGH" REGION.  
FT SITE 601 605 "KMSKS" REGION.  
FT BINDING 604 604 ATP (BY SIMILARITY).  
SQ SEQUENCE 820 AA; 93854 MW; 47DBB92B175B10B9 CRC64;

Query Match  
Best Local Similarity 8.5%; Score 78; DB 1; Length 820;  
Matches 46; Conservative 26; Mismatches 58; Indels 78; Gaps 11;

QY 13 EDLRCPETTSQA-----LPATTIEQASGEFGADEVOVFAP- 49  
DB 206 EDLDWPENVKQLQRNWICKSEALVRFEVNNERFLEVFTTR-----PDTIGGVSVFLWAPE 261

QY 50 ---ANAI PARSEAAAQVPVIGISQRVRMNSKEKD-----LGTLGY--VLGIT 92  
DB 262 HEVNRLLISENGREAVESVIRAQ-----SKSERDIRSETKVKTGVTGAKHPVTGAD 316

QY 93 MMV-----IIAIGAGIILGYSVKRGDKLKEOHDKVQEREMQRTTLPLSLAFTPTCBIV 147  
DB 317 IPIWISDYVLGVGSVGVGV-----PAHDER--DREP-----AEAFSLPIEVL 359  
QY 148 DKTVVVHTSQTPVDPQEGSTPLMGQAG 175  
DB 360 DKDECCIHGN-----HGDFLLDGLAG 380

RESULT 10  
SN\_HUMAN  
ID SN\_HUMAN STANDARD; PRT; 1709 AA.  
AC Q9BZ22; Q96DL4; Q9GZS5; Q9H1H7; Q9H7L7;  
DT 16-OCT-2001 (Rel. 40, Created)  
DT 16-OCT-2001 (Rel. 40, Last sequence update)  
DT 10-OCT-2003 (Rel. 42, Last annotation update)  
DE Sialoadhesin precursor (Sialic acid binding Ig-like lectin-1) (Siglec-1) (CD169 antigen).  
DE 1) (CD169 antigen).  
GN SN.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A. (ISOFORM 1), AND CHARACTERIZATION.  
RC TISSUE=Monocytes;  
RX MEDLINE=20575418; PubMed=11133773;  
RA Hartnell A., Steel J., Turley H., Jones M., Jackson D.G., Crocker P.R.;  
RA "Characterization of human sialoadhesin, a sialic acid binding receptor expressed by resident and inflammatory macrophage populations";  
RL Blood 97:288-296(2001).  
RN [2]  
RP SEQUENCE FROM N.A. (ISOFORMS 1 AND 2).  
RX MEDLINE=21638749; PubMed=11780052;  
RA Deloukas P., Matthews L.H., Ashurst J., Burton J., Gilbert J.G.R., Jones M., Stavridis G., Almeida J.P., Babbage A.K., Bagguley C.L., Bailey J., Barlow K.F., Bates K.N., Beard L.M., Beare D.M., Beasley O.P., Bird C.P., Blakey S.E., Bridgeman A.M., Brown A.J., Buck D., Burrill W.D., Butler A.P., Carder C., Carter N.P., Chapman J.C., Clapp M., Clark G., Clark L.N., Clark S.Y., Clee C.M., Clegg S., Cobley V.E., Collier R.E., Connor R.E., Corby N.R., Coulson A.G., Coville G.J., Deadman R., Dhali P.D., Dunn M., Ellington A.G., Frankland J.A., Fraser A., French L., Garner P., Graham D.V., Griffiths C., Griffiths M.N.D., Gwilliam R., Hall R.E., Hamond S., Harley J.L., Heath P.D., Ho S., Holden J.L., Howden P.J., Huckle E., Hunt A.R., Hunt S.E., Jekosch K., Johnson C.M., Johnson D., Kay M.P., Kimberley A.M., King A., Knights A., Laird G.K., Lawlor S., Lehaealaho M.H., Leversha M.A., Lloyd C., Lloyd D.M., Lovell J.D., Marsh V.L., Martin S.L., McConachie L.J., McLay K., McMurray A.A., Milne S.A., Mistry D., Moore M.J.F., Mullikin J.C., Nickerson T., Oliver K., Parker A., Patel R., Pearce T.A.V., Peck A.I., Phillimore B.J.C.T., Prathalingam S.R., Plumb R.M., Ramsay H., Rice C.M., Ross M.T., Scott C.E., Senra H.K., Shownkeen R., Sims S., Skuce C.D., Smith M.L., Soderlund C., Steward C.A., Sulston J.E., Swann R.M., Symons N., Taylor R., Tee L., Thomas D.W., Thorpe A., Tracey A., Trogan A.C., Vaudin M., Wall M., Wallis J.M., Whitehead S.L., Whittaker P., Willey D.L., Williams L., Williams S.A., Wilming L., Wray P.W., Hubbard T., Durbin R.M., Bentley D.R., Beck S., Rogers J.,  
RA "The DNA sequence and comparative analysis of human chromosome 20";  
RL Nature 414:865-871(2001).  
RN [3]  
RP SEQUENCE OF 733-1709 FROM N.A. (ISOFORMS 1 AND 2).  
RX MEDLINE=21082933; PubMed=11214971;  
RA Hattori A., Okumura K., Nagase T., Kikuno R., Hirosewa M., Ohara O.;  
RT "Characterization of long cDNA clones from human adult spleen";  
RL DNA Res. 7:357-366(2000).  
RN [4]  
RP SEQUENCE OF 1539-1709 FROM N.A. (ISOFORM 3).

TISSUE=Thymus;  
SUZUKI O., SASAKI N., AOTSUKA S., SHOJI T., ICHIHARA T., SHIOHATA N., MATSUMOTO K., HIRANO M., SANO S., NOMURA R., YOSHIKAWA Y., MATSUMURA Y., MORIYA S., CHIBA E., MONIYAMA H., KATAOKA S., KAEIYAMA S., SATOH N., MATSUNAWA H., TAKAHASHI E., KATAOKA R., KUGA N., KURODA A., SATOH I., KAMATA K., TAKAMI S., TERASHIMA Y., WATANABE M., SUGIYAMA T., IRIE R., OTSUKI T., SATO H., OTA T., WAKANABE A., ISHII S., YAMAMOTO J., ISONO Y., KAWAI-HIO Y., SAITO K., NISHIKAWA T., KIMURA K., YAMASHITA H., MATSUO K., NAKAMURA Y., SEKINE M., KIKUCHI H., KANDA K., WAGATSUMA M., MURAKAWA K., KANEHORI K., TAKAHASHI-FUJII A., OSHIMA A., SUGIYAMA A., KAWAKAMI B., SUZUKI Y., SUGANO S., NAGAHARI K., MASHUO Y., NAGAI K., ISOGAI T.;  
RT "NEDO human cDNA sequencing project";  
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.  
CC -!- FUNCTION: Macrophage-restricted adhesion molecule that mediates sialic-acid dependent binding to lymphocytes, including granulocytes, monocytes, natural killer cells, B-cells and CD8 T-cells. Preferentially binds to alpha2,3-linked sialic acid (By similarity). Binds to SPN/CD43 on T-cells (By similarity). May play a role in hemopoiesis.  
CC -!- SUBCELLULAR LOCATION: Type I membrane protein (isoform 1) and soluble (isoform 2).  
CC -!- ALTERNATIVE PRODUCTS:  
CC Event=Alternative splicing; Named isoforms=3;  
CC Comment=Additional isoforms seem to exist;  
CC Name=1;  
CC IsoId=Q9BZ22-1; Sequence=Displayed;  
CC Name=2;  
CC IsoId=Q9BZ22-2; Sequence=VSP\_002571;  
CC Name=3;  
CC IsoId=Q9BZ22-3; Sequence=VSP\_002572;  
CC Note=No experimental confirmation available;  
CC -!- TISSUE SPECIFICITY: Expressed by macrophages in various tissues. High levels are found in spleen, lymph node, perivascular macrophages in brain and lower levels in bone marrow, liver Kupffer cells and lamina propria of colon and lung. Also expressed by inflammatory macrophages in rheumatoid arthritis.  
CC -!- SIMILARITY: Belongs to the immunoglobulin superfamily. SIGLEC (sialic acid binding Ig-like lectin) family.  
CC -!- SIMILARITY: Contains 16 immunoglobulin-like V-type domain.  
CC -!- SIMILARITY: Contains 16 immunoglobulin-like C2-type domains.  
CC -!- DATABASE: NAFEP; PROW; NOTE=PROW 2:18-22(2001).  
CC WWW="http://www.ncbi.nlm.nih.gov/prow/guide/985165905\_g.htm".  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).  
CC EMBL; AF230073; AAC00757.1; -  
DR EMBL; AL109804; CAC17543.1; -  
DR EMBL; AL109804; CAC17542.1; -  
DR EMBL; AK024462; BAB15752.1; -  
DR EMBL; AK024459; BAB15749.1; -  
DR EMBL; AK024479; BAB15769.1; -  
DR EMBL; AK057560; BAB15727.1; -  
DR HSSP; Q62230; 1QFO.  
DR Genew; HGNC:11127; SN.  
DR MIM; 600751; -  
DR GO; GO:0016021; C:integral to membrane; NAS.  
DR GO; GO:0005529; F:sugar binding; NAS.  
DR GO; GO:0016337; P:cell-cell adhesion; NAS.  
DR GO; GO:0007160; P:cell-matrix adhesion; NAS.  
DR GO; GO:0006954; P:inflammatory response; NAS.  
DR InterPro; IPR007110; IG-like.  
DR InterPro; IPR003598; IG\_c2.  
DR Pfam; PF00047; Ig\_14\_c2.  
DR SMART; SM00408; IGC2; 4.  
DR PROSITE; PS50835; IG\_LIKE; 14.  
CC Cell adhesion; Lectin; Antigen; Transmembrane; Signal; Glycoprotein;





DR Pfam; PF00531; death; 1.  
 DR Pfam; PF00202; TNFR\_C6; 4.  
 DR SMART; SM00005; DEATH; 1.  
 DR SMART; SM00208; TNFR; 3.  
 DR PROSITE; PS00652; TNFR\_NGFR\_1; 3.  
 DR PROSITE; PS00652; TNFR\_NGFR\_2; 4.  
 DR PROSITE; PS00500; TNFR\_NGFR\_2; 4.  
 DR PROSITE; PS00017; DEATH\_DOMAIN; 1.  
 KW Receptor; Apoptosis; Neurogenesis; Transmembrane; Glycoprotein;  
 KW Repeat; Phosphorylation; Signal; 3D-structure.  
 FT SIGNAL 1 29  
 FT CHAIN 30 425

FT DOMAIN 30 251 TUMOR NECROSIS FACTOR RECEPTOR  
 FT TRANSMEM 252 273 SUPERFAMILY MEMBER 16.  
 FT DOMAIN 274 425 EXTRACELLULAR (POTENTIAL).  
 FT REPEAT 32 65 CYTOPLASMIC (POTENTIAL).  
 FT REPEAT 67 108 TNFR-CYS 1.  
 FT REPEAT 109 147 TNFR-CYS 2.  
 FT REPEAT 149 189 TNFR-CYS 3.  
 FT REPEAT 198 249 TNFR-CYS 4.  
 FT DOMAIN 33 44 DEATH.  
 FT DISULFID 33 44 SER/THR-RICH.  
 FT DISULFID 45 58 BY SIMILARITY.  
 FT DISULFID 48 65 BY SIMILARITY.  
 FT DISULFID 68 84 BY SIMILARITY.  
 FT DISULFID 87 100 BY SIMILARITY.  
 FT DISULFID 90 108 BY SIMILARITY.  
 FT DISULFID 110 123 BY SIMILARITY.  
 FT DISULFID 126 139 BY SIMILARITY.  
 FT DISULFID 129 147 BY SIMILARITY.  
 FT DISULFID 150 165 BY SIMILARITY.  
 FT DISULFID 168 181 BY SIMILARITY.  
 FT DISULFID 171 189 BY SIMILARITY.  
 FT CARBOHYD 61 61 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 71 71 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 SQ SEQUENCE 425 AA; 45432 MW; B2E152D94D3827F8 CRC64;

Query Match 8.4%; Score 76.5; DB 1; Length 425;  
 Best Local Similarity 24.4%; Pred. No. 15;  
 Matches 42; Conservative 23; Mismatches 90; Indels 17; Gaps 5;

QY 3 EAGVPEKPCEDLRCPETTSQALPAFTTETIQEASGEGGADVEQVFPAPNALPAREAAV 62  
 Db 175 ERQLRECTPWADAECEIPIGRWIPRSTPPGSDSTAPS\*QEPEV-PPEQDLVPSTVADNV 233  
 QY 63 QPVGIGISQRYNNSKEKIDGTIGVLTGMVILIIAGAILGVSYKRGKDLKEHQDQ 122  
 Db 234 TTWVGSSQPVVTR-----GTTDNLIPV-YCSILAAVVGLVAYAFRRNSCK-QNKQ 284  
 QY 123 KVCEREMQRTILPLSAFTNPTCEIVDEKTVVHTSTQTPVDPQEGSTPLMGQA 174  
 Db 285 GANSR-----PWNQPPPEGEKLSHSDSGISVDSQSLHDQQTHTQTASGQA 329

RESULT 12  
 GYG2 HUMAN  
 ID - GYG2 HUMAN STANDARD; PRT; 501 AA.  
 AC O15488; O15485; O15486; O15487; O15489; O15490;  
 DT 30-MAY-2000 (Rel. 39, Created)  
 DT 30-MAY-2000 (Rel. 39, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE Glycogenin-2 (EC 2.4.1.186) (GN-2) (GN2).  
 GN GYG2.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A., AND ALTERNATIVE SPLICING.  
 RC TISSUE=Liver;  
 RX MEDLINE=99010589; PubMed=9346895;  
 RA Mu J., Skurat A.V., Roach P.J.;  
 RT "Glycogenin-2, a novel self-glucosylating protein involved in liver

glycogen biosynthesis.";  
 J. Biol. Chem. 272:27589-27597(1997).  
 [2]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=20184741; PubMed=10721716;  
 RA Zhai L., Mu J., Zong H., DePaoli-Roach A.A., Roach P.J.;  
 RT "Structure and chromosomal localization of the human glycogenin-2 gene  
 GYG2.";  
 Gene 242:229-235(2000).  
 [3]  
 RP SEQUENCE FROM N.A. (ISOFORM BETA).  
 RC TISSUE=Skin;  
 RX MEDLINE=22388257; PubMed=12477932;  
 RA Sirauberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
 Klausner R.D., Collins F.S., Wagner L., Shennen C.M., Schuler G.D.,  
 Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,  
 Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,  
 Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,  
 Stapleton M., Soares M.B., Donald M.P., Casavant T.L., Scheetz T.E.,  
 Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,  
 Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,  
 Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
 Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,  
 Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
 Fehey J., Helton E., Kettman M., Madan A., Rodriguez S., Sanchez A.,  
 Whiting M., Madan A., Young A.C., Shcherchenko Y., Bouffard G.G.,  
 Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
 Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,  
 Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,  
 Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;  
 RT "Generation and initial analysis of more than 15,000 full-length  
 human and mouse cDNA sequences.";  
 Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).  
 [4]  
 RP CHARACTERIZATION, AND CARBOHYDRATE-LINKAGE SITE TYR-228.  
 RX MEDLINE=99074257; PubMed=9857012;  
 RA Mu J., Roach P.J.;  
 RT "Characterization of human glycogenin-2, a self-glucosylating  
 initiator of liver glycogen metabolism.";  
 J. Biol. Chem. 273:34850-34856(1998).  
 [5]  
 RP MASS SPECTROMETRY.  
 RC TISSUE=Breast cancer;  
 RX MEDLINE=21829512; PubMed=11840567;  
 RA Harris R.A., Yang A., Stein R.C., Lucy K., Brusten L., Herath A.,  
 Parekh R., Waterfield M.D., O'Hare M.J., Neville M.A., Page M.J.,  
 Zvelebil M.J.;  
 RT "Cluster analysis of an extensive human breast cancer cell line  
 protein expression map database.";  
 Proteomics 2:212-223(2002).  
 RL -!- FUNCTION: Self-glucosylates, via an inter-subunit mechanism, to  
 form an oligosaccharide primer that serves as substrate for  
 glycogen synthase.  
 CC -!- CATALYTIC ACTIVITY: UDP-glucose + glycogenin = UDP +  
 glucosylglycogenin.  
 CC -!- COFACTOR: Self-glucosylation is dependent on the presence of  
 divalent metal ions of which manganese ion is the most effective.  
 CC -!- PATHWAY: Glycogen biosynthesis.  
 CC -!- SUBUNIT: Homodimer, tightly complexed to glycogen synthase.  
 CC -!- ALTERNATIVE PRODUCTS:  
 CC Event=Alternative splicing; Named isoforms=6;  
 CC Comment=Additional isoforms seem to exist;  
 CC Name=Alpha;  
 CC IsoId=O15488-1; Sequence=Displayed;  
 CC Name=Beta;  
 CC IsoId=O15488-2; Sequence=VSP\_001770;  
 CC Name=Gamma;  
 CC IsoId=O15488-3; Sequence=VSP\_001771;  
 CC Name=Delta;  
 CC IsoId=O15488-4; Sequence=VSP\_001772;  
 CC Name=Epsilon;  
 CC IsoId=O15488-5; Sequence=VSP\_001773;  
 CC Name=Zeta;



CC -i- SUBUNIT: Homomultimer and heteromultimer with other K-C1  
CC cotransporters (By similarity).  
CC -i- SUBCELLULAR LOCATION: Integral membrane protein.  
CC -i- TISSUE SPECIFICITY: Brain-specific. Detected in neuronal cells.  
CC -i- MISCELLANEOUS: Inhibited by furosemide and bumetanide.  
CC -i- SIMILARITY: BELONGS TO THE SLC12A FAMILY OF TRANSPORTERS.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; AF208159; AAC43493.1; -  
DR EMBL; AL162458; CAC10460.1; -  
DR EMBL; AB033002; BAB86490.1; -  
DR Genew; HGNC:13818; SLC12A5.  
DR MIM; 606726; -  
DR GO; GO:0016021; C:integral to membrane; NAS.  
DR GO; GO:0015379; F:potassium:chloride symporter activity; NAS.  
DR GO; GO:0006873; P:cell ion homeostasis; NAS.  
DR GO; GO:0006810; P:transport; NAS.  
DR GO; GO:0006810; P:transport; NAS.  
DR InterPro; IPR002293; AA/rel\_permease1.  
DR InterPro; IPR004842; KCL cotranspt.  
DR InterPro; IPR000076; KCL cotranspt.  
DR InterPro; IPR004841; Permease region.  
DR Pfam; PF00324; aa\_permeases; 1.  
DR PRINTS; PR01081; KCLTRANSPORT.  
DR TIGRFAMs; TIGR00930; 2a30; 1.  
DR Transmembrane. 1  
KW Transport; ion transport; Symport; Potassium; Potassium transport;  
KW Transmembrane. 1 110 CYTOPLASMIC (POTENTIAL).  
FT DOMAIN 111 131 POTENTIAL.  
FT TRANSMEM 133 153 POTENTIAL.  
FT TRANSMEM 154 171 POTENTIAL.  
FT DOMAIN 172 192 POTENTIAL.  
FT TRANSMEM 194 214 POTENTIAL.  
FT DOMAIN 215 231 POTENTIAL.  
FT TRANSMEM 232 252 POTENTIAL.  
FT TRANSMEM 255 275 POTENTIAL.  
FT DOMAIN 276 395 POTENTIAL.  
FT TRANSMEM 396 416 POTENTIAL.  
FT TRANSMEM 436 456 POTENTIAL.  
FT DOMAIN 457 473 CYTOPLASMIC (POTENTIAL).  
FT TRANSMEM 474 494 POTENTIAL.  
FT TRANSMEM 547 567 POTENTIAL.  
FT DOMAIN 568 607 POTENTIAL.  
FT TRANSMEM 608 628 POTENTIAL.  
FT TRANSMEM 825 845 POTENTIAL.  
FT DOMAIN 846 1116 CYTOPLASMIC (POTENTIAL).  
FT CARBOHYD 419 419 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 810 810 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CONFLICT 2 2 P -> L (IN REF. 2).  
FT CONFLICT 1077 1077 P -> L (IN REF. 3).  
SQ SEQUENCE 1116 AA; 123495 MW; 7C857D58DDAB4264 CRC64;  
  
Query Match 8.4%; Score 76.5; DB 1; Length 1116;  
Best Local Similarity 26.3%; Pred. No. 44;  
Matches 31; Conservative 16; Mismatches 44; Indels 27; Gaps 4;  
  
QY 79 KDLGTGLVGLITMIIIAIGAGIILGYSY-----KRGDKLKEQH-DQKVCEREMQ 130  
DB 872 KDLTTLPLHLRITAEEVEMHESDISAVTYEKLVMQORSQILKOMHLTKNEREREIQ 931  
QY 131 RIT-----LPLSA---FTNPTCIVDEKTVVWHTSTQPDVDPQEGTTP 169  
DB 932 SITDESRSIRKPNANTRLNVPETAGDSEKPEEVLQHDQSPSCPSSSP 989  
  
RESULT 14  
SPKD\_SYNY3

ID SPKD\_SYNY3 STANDARD; PRT; 505 AA.  
AC PS4735;  
DT 01-OCT-1996 (Rel. 34, Created)  
DT 01-OCT-1996 (Rel. 34, Last sequence update)  
DT 15-MAR-2004 (Rel. 43, Last annotation update)  
DE Serine/threonine-protein kinase D (EC 2.7.1.37).  
OS SPKD OR SLL0776.  
GN Synechocystis sp. (strain PCC 6803).  
OC Bacteria; Cyanobacteria; Chroococcales; Synechocystis.  
OX NCBI\_TaxID=1148;  
RN [1]  
RP SEQUENCE FROM N.A., AND CHARACTERIZATION.  
RP MEDLINE=22158630; PubMed=12168951;  
RA Kamei A., Yuasa T., Gang X., Ikeuchi M.;  
RA "Biochemical examination of the potential eukaryotic-type protein  
RT kinase genes in the complete genome of the unicellular Cyanobacterium  
RT synechocystis sp. PCC 6803.";  
RL DNA Res. 9:71-78(2002).  
RN [2]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=96127529; PubMed=8590279;  
RA Kaneko T., Tanaka A., Sato S., Kotani H., Sazuka T., Miyajima N.,  
RA Sugitara M., Tabata S.;  
RT "Sequence analysis of the genome of the unicellular cyanobacterium  
RT Synechocystis sp. strain PCC6803. I. Sequence features in the 1 Mb  
RT region from map positions 64% to 92% of the genome.";  
RL DNA Res. 2:153-166(1995).  
CC -i- CATALYTIC ACTIVITY: ATP + a protein = ADP + a phosphoprotein.  
CC -i- SIMILARITY: Belongs to the Ser/Thr family of protein kinases.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; AB046600; BAB17036.1; -  
DR EMBL; D64005; BAA10726.1; -  
DR PIR; S77034; S77034.  
DR InterPro; IPR000719; Prot\_kinase.  
DR InterPro; IPR008271; Ser\_Thr\_kin\_AS.  
DR InterPro; IPR003646; SH3\_bac.  
DR Pfam; PF00069; pkinase; 1.  
DR ProDom; PD000001; Prot\_kinase; 1.  
DR SMART; SM00287; SH3b; 1.  
DR PROSITE; PS00107; PROTEIN\_KINASE\_ATP; 1.  
DR PROSITE; PS00108; PROTEIN\_KINASE\_ST; 1.  
DR PROSITE; PS50011; PROTEIN\_KINASE\_DOM; 1.  
KW Transferase; Serine/threonine-protein kinase; ATP-binding;  
KW Complete proteome.  
FT DOMAIN 9 271 PROTEIN KINASE.  
FT NP\_BIND 15 23 ATP (BY SIMILARITY).  
FT BINDING 40 40 ATP (BY SIMILARITY).  
FT ACT\_SITE 136 136 BY SIMILARITY.  
SQ SEQUENCE 505 AA; 55213 MW; C4F12A1886C4D51C CRC64;  
  
Query Match 8.3%; Score 76; DB 1; Length 505;  
Best Local Similarity 27.1%; Pred. No. 20;  
Matches 35; Conservative 17; Mismatches 69; Indels 8; Gaps 4;  
  
QY 8 EKRCEDLRCEPTTSQALPAFTTEIQEASGSGADEVQVPAPANA-LPARSEAAAVQPV- 65  
DB 250 EMPSRRYSSAEAMVQALHSL---ISSGAEPALPMETVRVAPSNFELVTSSTKTATVV 306  
QY 66'--IGISQVRVRNNSKEKDLGLTGLVGLITMIIIAIGAGIILGYSYKRGDKLKEHQK 123  
DB 307 KEVGNHNNYNNNGSKSIATLLTVL-IGIIVTAGLGGGFIITQIQKEAEARAAQAEKE 365  
QY 124 VCEREMQRI 132  
DB 366 KOAEQKRI 374

RESULT 15	ID	TRN8 HUMAN	STANDARD;	PRT;	595 AA.
DR	AC	P28908;			
DT	DT	01-DEC-1992 (Rel. 24, Created)			
DT	DT	01-DEC-1992 (Rel. 24, Last sequence update)			
DT	DT	10-OCT-2003 (Rel. 42, Last annotation update)			
DE	DE	Tumor necrosis factor receptor superfamily member 8 precursor (CD30L receptor) (lymphocyte activation antigen CD30) (Ki-1 antigen).			
GN	GN	TNFRSF8 OR CD30.			
OC	OC	Homo sapiens (Human).			
OC	OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			
OC	OC	NCBI_TaxID:9606;			
RP	RP	[1] SEQUENCE FROM N.A. (ISOFORM LONG).			
RP	RP	TISSUE=Lymphoid;			
RP	RP	MEDLINE=92154659; PubMed=1310894;			
RA	RA	Duerkop H., Latza U., Hummel F., Seed B., Stein H.;			
RT	RT	"Molecular cloning and expression of a new member of the nerve growth factor receptor family that is characteristic for Hodgkin's disease";			
RT	RT	Cell 68:421-427(1992).			
RT	RT	[2]			
RP	RP	SEQUENCE FROM N.A. (ISOFORM LONG).			
RP	RP	MEDLINE=95089787; PubMed=7527901;			
RA	RA	Jung W., Krueger S., Renner C., Gause A., Sahin U., Trumper L.,			
RA	RA	Pfreundschuh M.;			
RT	RT	"Opposite effects of the CD30 ligand are not due to CD30 mutations: results from cDNA cloning and sequence comparison of the CD30 antigen from different sources";			
RT	RT	Mol. Immunol. 31:1329-1334(1994).			
RT	RT	[3]			
RP	RP	SEQUENCE FROM N.A. (ISOFORM SHORT).			
RP	RP	MEDLINE=96437016; PubMed=8839832;			
RA	RA	Horie R., Ito K., Tatewaki M., Nagai M., Aizawa S.,			
RA	RA	Higashihara M., Ishida T., Inoue J.-I., Takizawa H., Watanabe T.;			
RT	RT	"A variant CD30 protein lacking extracellular and transmembrane domain is induced in HL-60 by tetradecanoylphorbol acetate and is expressed in alveolar macrophages";			
RT	RT	Blood 88:2422-2432(1996).			
RT	RT	[4]			
RP	RP	INTERACTION WITH TRAF1 AND TRAF2.			
RP	RP	MEDLINE=96195221; PubMed=8627180;			
RA	RA	Lee S.Y., Park C.G., Choi Y.;			
RT	RT	"T cell receptor-dependent cell death of T cell hybridomas mediated by the CD30 cytoplasmic domain in association with tumor necrosis factor receptor-associated factors";			
RT	RT	J. Exp. Med. 183:669-674(1996).			
RT	RT	[5]			
RP	RP	INTERACTION WITH TRAF3.			
RP	RP	MEDLINE=97312455; PubMed=9168896;			
RA	RA	Boucher L.M., Marengere L.E., Lu Y., Thukral S., Mak T.W.;			
RT	RT	"Binding sites of cytoplasmic effectors TRAF1, 2, and 3 on CD30 and other members of the TNF receptor superfamily";			
RT	RT	Biochem. Biophys. Res. Commun. 233:592-600(1997).			
RT	RT	[6]			
RP	RP	INTERACTION WITH TRAF5.			
RP	RP	MEDLINE=98172745; PubMed=9511754;			
RA	RA	Mizushima S.-I., Fujita M., Ishida T., Azuma S., Kato K., Hirai M.,			
RA	RA	Otsuka M., Yamamoto T., Inoue J.-I.;			
RT	RT	"Cloning and characterization of a cDNA encoding the human homolog of tumor necrosis factor receptor-associated factor 5 (TRAF5).";			
RT	RT	Gene 207:135-140(1998).			
RT	RT	[7]			
RP	RP	INTERACTION WITH TRAF6.			
RP	RP	MEDLINE=97152965; PubMed=8998989;			
RA	RA	Aizawa S., Nakano H., Ishida T., Horie R., Nagai M., Ito K.,			
RA	RA	Yagita H., Okumura K., Inoue J.-I., Watanabe T.;			
RT	RT	"Tumor necrosis factor receptor-associated factor (TRAF) 5 and TRAF2 are involved in CD30-mediated NF-kappaB activation";			
RT	RT	(POTENTIAL).";			

```
FT CARBOHYD 276 276 N-LINKED (GLCNAC... ) (POTENTIAL).
SQ SEQUENCE 595 AA; 63747 MW; 7A407CC78A6E0BC8 CRC64;

Query Match 8.2%; Score 75; DB 1; Length 595;
Best Local Similarity 23.1%; Pred. No. 30;
Matches 43; Conservative 31; Mismatches 70; Indels 42; Gaps 10;

QY 13 EDLRCPETTSQALPAPT-----TEIQFASGPGADEVQVFAPANALPARSEAAAVOPVIGI 68
Db :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|:
315 QDMAEKDTTTFEAPPLGTQPCDNCPTPENGAPAST-----SPTQSLLVDSQASKTLP-IPT 368

QY 69 SQVRVNSKEKKDLGTGLGYVLGITMMVIIAIGAGIILGYSYKRGKDLKEQHDQKVCERE 128
Db :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|:
369 SAPVALSSTGKPVLDG-GPVLFWVLVLVVVVGSSAFL-----LCHREACKR 415

QY 129 MQR---ITLPLSAFTNPTCEIIVDEKTVVHTSQT-----PVDPOEG--STPLMGQ 173
Db :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|:
416 IRQKLHLCYFVQT-SQPKLELVDSRP---RRSSTQLRSGASVTEPVAEERGLMSQFLMET 471

QY 174 AGTPGA 179
Db :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|:
472 CHSVGA 477
```

Search completed: March 17, 2004, 07:04:28  
Job time : 10.2275 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 17, 2004, 06:59:03 ; Search time 32.7645 Seconds  
(without alignments)  
1723.750 Million cell updates/sec

Title: US-10-057-951-2\_COPY\_85\_263

Perfect score: 916

Sequence: 1 SGBAYPEKPCEDLRCPET.....PVDPPGSGTFLMGQAGTGA 179

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- 1: sp\_archaea.\*
- 2: sp\_bacteria.\*
- 3: sp\_fungi.\*
- 4: sp\_human.\*
- 5: sp\_invertebrate.\*
- 6: sp\_mammal.\*
- 7: sp\_rhiz.\*
- 8: sp\_organelle.\*
- 9: sp\_phage.\*
- 10: sp\_plant.\*
- 11: sp\_rodent.\*
- 12: sp\_virus.\*
- 13: sp\_vertebrate.\*
- 14: sp\_unclassified.\*
- 15: sp\_rviro.\*
- 16: sp\_bacterioph.\*
- 17: sp\_archaeop.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	916	100.0	263	4 Q96FE7	Q96fe7 homo sapien
2	913	99.7	263	4 Q00318	Q00318 homo sapien
3	902	98.5	263	4 Q8NCU9	Q8ncj9 homo sapien
4	732.5	80.0	264	11 Q7TMJ8	Q7tmj8 mus musculus
5	728.5	79.5	264	11 Q811Z2	Q811z2 mus musculus
6	562	61.4	234	4 Q86YW2	Q86yw2 homo sapien
7	487	53.2	263	13 Q7SKB3	Q7skb3 brachydanio
8	438.5	47.9	213	11 Q811Z3	Q811z3 mus musculus
9	33.5	10.2	569	10 Q9M261	Q9m261 arabidopsis
10	87.5	9.6	759	5 Q20429	Q20429 caenorhabdi
11	86	9.4	290	5 Q9BL62	Q9bl62 caenorhabdi
12	86	9.4	479	5 Q9GZG5	Q9gzg5 caenorhabdi
13	85.5	9.3	1114	11 Q7TQC9	Q7tqc9 mus musculus
14	85.5	9.3	1164	11 Q80TIS	Q80tis mus musculus
15	84.5	9.2	607	17 Q8U3P9	Q8u3p9 pyrococcus
16	84.5	9.2	733	16 Q879Z2	Q879z2 xyliella fas

17	83.5	9.1	517	16 Q93J57	Q93j57 streptomyce
18	83.5	9.1	733	16 Q9PA31	Q9pa31 xyliella fas
19	83	9.1	1553	5 Q95Q77	Q95q77 caenorhabdi
20	82.5	9.0	503	11 Q8C165	Q8c165 mus musculus
21	82	9.0	774	10 Q81I28	Q81i28 oryza sativ
22	81.5	8.9	535	10 Q8M4W9	Q8m4w9 solanum tub
23	81.5	8.9	715	10 Q9W839	Q9w839 arabidopsis
24	81.5	8.9	1352	5 Q8MTN9	Q8mtn9 caenorhabdi
25	81	8.8	500	5 Q81LQ2	Q81lq2 plasmodium
26	80.5	8.8	361	16 Q7UH23	Q7uh23 rhodopirell
27	80.5	8.8	437	16 Q88HQ0	Q88hq0 pseudomonas
28	80.5	8.8	483	16 Q89AB1	Q89ae1 buchnera ap
29	80.5	8.8	803	16 Q25153	Q25153 helicobacte
30	80.5	8.8	918	16 Q8XRT1	Q8xrt1 ralstonia s
31	80.5	8.8	1852	3 Q9C2H4	Q9c2h4 neurospora
32	80.5	8.8	2358	16 Q9LIV8	Q9liv8 streptomyce
33	80	8.7	618	2 Q9F5E5	Q9f5e5 agrobacteri
34	79.5	8.7	452	16 Q8CFQ9	Q8cfq9 staphylococ
35	79.5	8.7	664	16 Q88U15	Q88u15 lactobacill
36	79.5	8.7	770	2 Q9AQR9	Q9agr9 bacillus th
37	79	8.6	717	16 Q886D3	Q886d3 pseudomonas
38	79	8.6	960	16 Q92KP5	Q92kp5 rhizobium m
39	79	8.6	1091	16 Q88LM4	Q88lm4 pseudomonas
40	78.5	8.6	457	16 Q8UHM4	Q8uhm4 agrobacteri
41	78.5	8.6	478	5 Q25684	Q25684 plasmodium
42	78.5	8.6	552	5 Q9VE62	Q9ve62 drosophila
43	78.5	8.6	2972	5 P90891	P90891 caenorhabdi
44	78	8.5	224	16 Q8Z5M1	Q8z5m1 salmonella
45	78	8.5	336	17 Q8ZUF0	Q8zuf0 pyrobaculum

ALIGNMENTS

RESULT 1  
Q96FE7  
ID Q96FE7 PRELIMINARY; PRT; 263 AA.  
AC Q96FE7, 2001 (Tremblrel. 19, Created)  
DT 01-DEC-2001 (Tremblrel. 19, Last sequence update)  
DT 01-DEC-2001 (Tremblrel. 19, Last sequence update)  
DT 01-JUN-2003 (Tremblrel. 24, Last annotation update)  
DE Hypothetical protein (HGFL(L) protein).  
GN HGFL.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Brain;  
RA Strausberg R.;  
RL Submitted (JUL-2001) to the EMBL/GenBank/DBSJ databases.  
[2]  
RP SEQUENCE FROM N.A.  
RA Chiang H., Chang M.;  
RL Submitted (JUL-2002) to the EMBL/GenBank/DBSJ databases.  
CC -!- SIMILARITY: CONTAINS 1 KRINGLE DOMAIN.  
DR EMBL; BC011049; AAH11049.1; -!  
DR EMBL; AF528080; AAC33763.1; -!  
DR InterPro; IPR000001; Kringle.  
DR Pfam; PF00051; kringle; 1.  
DR PRINTS; PR00018; KRINGLE.  
DR ProDom; PD000395; Kringle; 1.  
DR SMART; SM00130; KR; 1.  
DR PROSITE; PS00021; KRINGLE\_1; 1.  
DR PROSITE; PS00070; KRINGLE\_2; 1.  
KW Hypothetical protein; Glycoprotein; Kringle.  
SQ SEQUENCE 263 AA; 28234 MW; 197C9EE888FA242 CRC64;

Query Match 100.0%; Score 916; DB 4; Length 263;  
Best Local Similarity 100.0%; Pred. No. 8e-80;  
Matches 179; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY 1 SGEAGVPEKPCEDLRCPETTSQALPAFTTEIQEASGPGADEVQVFAPANALPARSEAA 60
DB 85 SGEAGVPEKPCEDLRCPETTSQALPAFTTEIQEASGPGADEVQVFAPANALPARSEAA 144
QY 61 AVQPVIGISQVRMNSKEKDLGTLGYVLGTTMVIITAIAGAILGYSYKRGKDLKEQH 120
DB 145 AVQPVIGISQVRMNSKEKDLGTLGYVLGTTMVIITAIAGAILGYSYKRGKDLKEQH 204
QY 121 DQKVCEREMQRTITPLSAFTNPTCEIVDEKTVVHTSQTPVDPQEGSTPLMGQAGTPGA 179
DB 205 DQKVCEREMQRTITPLSAFTNPTCEIVDEKTVVHTSQTPVDPQEGSTPLMGQAGTPGA 263

RESULT 2
O00318 PRELIMINARY; PRT; 263 AA.
AC O00318;
DT 01-JUL-1997 (T-EMBLrel. 04, Created)
DT 01-JUL-1997 (T-EMBLrel. 04, Last sequence update)
DT 01-OCT-2003 (T-EMBLrel. 25, Last annotation update)
DE WUGSC:DJ515N1.2 protein.
GN WUGSC:DJ515N1.2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
CX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Du Z., Scheet P., Harper M.;
RT "The sequence of H. sapiens PAC clone RP3-515N1."
RL Submitted (MAY-1997) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA Waterston R.;
RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.
CC -!- SIMILARITY: CONTAINS 1 KRINGLE DOMAIN.
DB EMBL; AC002073; AAB54054.1; -.
DR HSP; P00749; 1KDU.
DR InterPro: IPR000001; Kringle.
DR Pfam: PF00051; kringle.1.
DR PRINTS; PR00018; KRINGLE.
DR ProDom; PD000395; Kringle; 1.
DR SMART; SM00130; KR; 1.
DR PROSITE; PS00021; KRINGLE_1; 1.
DR PROSITE; PS00070; KRINGLE_2; 1.
KW Glycoprotein; Kringle.
SQ SEQUENCE 263 AA; 28248 MW; 197C3BEE8E54A242 CRC64;

Query Match 99.7%; Score 91.3; DB 4; Length 263;
Best Local Similarity 99.4%; Pred. No. 1.5e-79;
Matches 178; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 SGEAGVPEKPCEDLRCPETTSQALPAFTTEIQEASGPGADEVQVFAPANALPARSEAA 60
DB 85 SGEAGVPEKPCEDLRCPETTSQALPAFTTEIQEASGPGADEVQVFAPANALPARSEAA 144
QY 61 AVQPVIGISQVRMNSKEKDLGTLGYVLGTTMVIITAIAGAILGYSYKRGKDLKEQH 120
DB 145 AVQPVIGISQVRMNSKEKDLGTLGYVLGTTMVIITAIAGAILGYSYKRGKDLKEQH 204
QY 121 DQKVCEREMQRTITPLSAFTNPTCEIVDEKTVVHTSQTPVDPQEGSTPLMGQAGTPGA 179
DB 205 DQKVCEREMQRTITPLSAFTNPTCEIVDEKTVVHTSQTPVDPQEGSTPLMGQAGTPGA 263

RESULT 3
Q8NCJ9 PRELIMINARY; PRT; 263 AA.
AC Q8NCJ9;
DT 01-OCT-2002 (T-EMBLrel. 22, Created)
DT 01-OCT-2002 (T-EMBLrel. 22, Last sequence update)
DT 01-JUN-2003 (T-EMBLrel. 24, Last annotation update)
DE Hypothetical protein FLJ90207.

```

```

OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
CX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Isogai T., Ota T., Nishikawa T., Hayashi K., Otsuki T., Sugiyama T.,
RA Suzuki Y., Nagai K., Sugano S., Ishii S., Kawai-Hio Y., Saito K.,
RA Yamamoto Y., Wakamatsu A., Nakamura Y., Kojima S., Nagahara K.,
RA Masuho J., Ono T., Okano K., Yoshikawa Y., Aotsuka S., Sasaki N.,
RA Hattori A., Okumura K., Iwayanagi T., Ninomiya K.;
RT "NEDO human cDNA sequencing project."
RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
CC -!- SIMILARITY: CONTAINS 1 KRINGLE DOMAIN.
DB EMBL; AK074688; BAC11140.1; -.
DR InterPro; IPR000001; Kringle.
DR Pfam; PF00051; kringle.1.
DR PRINTS; PR00018; KRINGLE.
DR ProDom; PD000395; Kringle; 1.
DR SMART; SM00130; KR; 1.
DR PROSITE; PS00021; KRINGLE_1; 1.
DR PROSITE; PS00070; KRINGLE_2; 1.
KW Hypothetical protein; Glycoprotein; Kringle.
SQ SEQUENCE 263 AA; 28104 MW; 73A9294D5142C6C3 CRC64;

Query Match 98.5%; Score 902; DB 4; Length 263;
Best Local Similarity 98.9%; Pred. No. 1.8e-78;
Matches 177; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 SGEAGVPEKPCEDLRCPETTSQALPAFTTEIQEASGPGADEVQVFAPANALPARSEAA 60
DB 85 SGEAGVPEKPCEDLRCPETTSQALPAFTTEIQEASGPGADEVQVFAPANALPARSEAA 144
QY 61 AVQPVIGISQVRMNSKEKDLGTLGYVLGTTMVIITAIAGAILGYSYKRGKDLKEQH 120
DB 145 AVQPVIGISQVRMNSKEKDLGTLGYVLGTTMVIITAIAGAILGYSYKRGKDLKEQH 204
QY 121 DQKVCEREMQRTITPLSAFTNPTCEIVDEKTVVHTSQTPVDPQEGSTPLMGQAGTPGA 179
DB 205 DQKVCEREMQRTITPLSAFTNPTCEIVDEKTVVHTSQTPVDPQEGSTPLMGQAGTPGA 263

RESULT 4
Q7TMJ8 PRELIMINARY; PRT; 264 AA.
AC Q7TMJ8;
DT 01-OCT-2003 (T-EMBLrel. 25, Created)
DT 01-OCT-2003 (T-EMBLrel. 25, Last sequence update)
DT 01-OCT-2003 (T-EMBLrel. 25, Last annotation update)
DE Hypothetical protein.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
CX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN=C57BL/6J; TISSUE=Breast tumor;
RA MEDLINE=22388257; PubMed=12477932;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeng B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S., Loquellano N.A., Peters G.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Kettaman M., Madan A., Rodrigues S., Sanchez A.,
RA Whitting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,

```

RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,  
RA Arzayuskii M.I., Skalska U., Smalish D.E., Schnerch A., Schein J.E.,  
RA Jones S.J., Marra M.A.;  
RT "Generation and initial analysis of more than 15,000 full-length human  
RT and mouse cDNA sequences.";  
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=C57BL/6J; TISSUE=Breast tumor;  
RA Strausberg R.;  
RL Submitted (AUG-2003) to the EMBL/GenBank/DBJ databases.  
DR EMBL; BC055920; AAH55920.1; -;  
KW Hypothetical protein.  
SQ SEQUENCE 264 AA; 28567 MW; 833EA578FEB763A4 CRC64;  
  
Query Match 80.0%; Score 732.5; DB 11; Length 264;  
Best Local Similarity 79.4%; Pred. No. 3.5e-62;  
Matches 143; Conservative 13; Mismatches 23; Indels 1; Gaps 1;  
  
QY 1 SGEAGVPEKPCEDLRCPETTSQA-LPAFTTEIOEASGPGADVQVFPANALPARSEA 59  
DB EMBL; AF528079; AA033762.1; -;  
85 SSETGVPEKPCEDVSCPTTSQAPPPSSAMELEKSGAPGDKAQPFPANALPARSEA 144  
QY 60 AAVQPVIGISQVRVNSKEKDLGTLGYVLGTTMWIIAIGAGIILGYSGKDLKEQ 119  
DB 145 AEVQPVIGISQVRVNSKEKDLGTLGYVLGTTMWIIAIGAGIILGYSGKDLKEQ 204  
QY 120 HDQKVCEREMORITPLSAFTNPTCEIVDEKTVVHTSQTPVDPQEGSTPLMGQAGTPGA 179  
DB 205 HEKKACEREMORITPLSAFTNPTCETVDENTIIHNSNTPADVQEGSTLLTGQAGTPGA 264  
  
RESULT 5  
Q81122 PRELIMINARY; PRT; 264 AA.  
ID Q81122  
AC Q81122  
DT 01-JUN-2003 (TREMBlrel. 24, Created)  
DT 01-JUN-2003 (TREMBlrel. 24, Last sequence update)  
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)  
DE HGFL(L) protein.  
GN HGFL  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=BALB/c;  
RA Chiang H., Chang M.;  
RL Submitted (JUL-2002) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AF528081; AA033764.1; -;  
DR InterPro; IPR000001; Kringle.  
DR Pfam; PF00051; Kringle; 1.  
DR PRINTS; PR00018; KRINGLE.  
DR SMART; SM00130; KR; 1.  
DR PROSITE; PS00021; KRINGLE\_1; 1.  
DR PROSITE; PS50070; KRINGLE\_2; 1.  
SQ SEQUENCE 264 AA; 28595 MW; 833EA578FEB50E34 CRC64;  
  
Query Match 79.5%; Score 728.5; DB 11; Length 264;  
Best Local Similarity 78.9%; Pred. No. 8.4e-62;  
Matches 142; Conservative 14; Mismatches 23; Indels 1; Gaps 1;  
  
QY 1 SGEAGVPEKPCEDLRCPETTSQA-LPAFTTEIOEASGPGADVQVFPANALPARSEA 59  
DB 85 SSETGVPEKPCEDVSCPTTSQAPPPSSAMELEKSGAPGDKAQPFPANALPARSEA 144  
QY 60 AAVQPVIGISQVRVNSKEKDLGTLGYVLGTTMWIIAIGAGIILGYSGKDLKEQ 119  
DB 145 AEVQPVIGISQVRVNSKEKDLGTLGYVLGTTMWIIAIGAGIILGYSGKDLKEQ 204  
QY 120 HDQKVCEREMORITPLSAFTNPTCEIVDEKTVVHTSQTPVDPQEGSTPLMGQAGTPGA 179

Db 205 HEKKACEREMORITPLSAFTNPTCETVDENTIIHNSNTPADVQEGSTLLTGQAGTPGA 264  
  
RESULT 6  
Q86YW2 PRELIMINARY; PRT; 234 AA.  
ID Q86YW2  
AC Q86YW2  
DT 01-JUN-2003 (TREMBlrel. 24, Created)  
DT 01-JUN-2003 (TREMBlrel. 24, Last sequence update)  
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)  
DE HGFL(S) protein.  
GN HGFL  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Chang H., Chang M.;  
RL Submitted (JUL-2002) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AF528079; AA033762.1; -;  
DR InterPro; IPR000001; Kringle.  
DR Pfam; PF00051; Kringle; 1.  
DR PRINTS; PR00018; KRINGLE.  
DR SMART; SM00130; KR; 1.  
DR PROSITE; PS00021; KRINGLE\_1; 1.  
DR PROSITE; PS50070; KRINGLE\_2; 1.  
SQ SEQUENCE 234 AA; 25320 MW; C78F64DBD1B8DC0D CRC64;  
  
Query Match 61.4%; Score 562; DB 4; Length 234;  
Best Local Similarity 100.0%; Pred. No. 7.3e-46;  
Matches 112; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 1 SGEAGVPEKPCEDLRCPETTSQA-LPAFTTEIOEASGPGADVQVFPANALPARSEA 60  
DB 85 SGEAGVPEKPCEDLRCPETTSQA-LPAFTTEIOEASGPGADVQVFPANALPARSEA 144  
QY 61 AVQPVIGISQVRVNSKEKDLGTLGYVLGTTMWIIAIGAGIILGYSGYKR 112  
DB 145 AVQPVIGISQVRVNSKEKDLGTLGYVLGTTMWIIAIGAGIILGYSGYKR 196  
  
RESULT 7  
Q7SXB3 PRELIMINARY; PRT; 263 AA.  
ID Q7SXB3  
AC Q7SXB3  
DT 01-OCT-2003 (TREMBlrel. 25, Created)  
DT 01-OCT-2003 (TREMBlrel. 25, Last sequence update)  
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)  
DE Hypothetical protein.  
OS Brachydanio rerio (Zebrafish) (Danio rerio).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;  
OC Cyprinidae; Danio.  
OX NCBI\_TaxID=7955;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=AB; TISSUE=Body;  
RX MEDLINE=23289257; PubMed=12477932;  
RA Strausberg R.B., Feingold E.A., Grouse L.H., Derge J.G.,  
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,  
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,  
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Heide F.,  
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,  
RA Stachenko M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,  
RA Brownstein M.J., Ustin T.B., Toshiyuki S., Carninci P., Prange C.,  
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,  
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,  
RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
RA Fahey J., Helton E., Kettman M., Madan A., Rodriguez S., Sanchez A.,  
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.



Db 145 A5VQPVIGISQVRNMSKEKDLGLTGLVYVIGITMMVILLIAGIAGIIVGYTKR 197

RESULT 9

Q9M261 Q9M261 PRELIMINARY; PRT; 569 AA.

AC Q9M261; 01-OCT-2000 (TrEMBLrel. 15, Created)

DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)

DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)

DE Hypothetical protein.

GN F7M19.10.

OS Arabidopsis thaliana (Mouse-ear cress).

OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;

OC euroids II; Brassicales; Brassicaceae; Arabidopsis.

OX NCBI\_TaxID=3702;

RN [1]

RP SEQUENCE FROM N.A.

RA Nakamura G., Fartmann B., Dauner D., Sterr W., Holland R.,

RA Weichselgartner M., Mewes H.W., Rudd S., Lemcke K., Mayer K.F.X.,

RA Quettier F., Salanoubat M.,

RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.

RP [2]

RP SEQUENCE FROM N.A.

RA EU Arabidopsis sequencing project;

RL Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.

DR EMBL; ALI38643; CAB86471.1; -

DR F1R; T47358; T47358.

KW Hypothetical protein.

SQ SEQUENCE 569 AA; 63537 MW; B775942AC7C565F5 CRC64;

Query Match 10.2%; Score 93.5; DB 10; Length 569;

Best Local Similarity 32.1%; Pred. No. 2.4;

Matches 27; Conservative 10; Mismatches 36; Indels 11; Gaps 1;

QY 8 EKRPCELRCPETTSQALPAFTTETIQEASGPGADVQVFAPAN-----ALPAR 56

DB 301 KGRAAEERVVESGPAEPVPEVAAPVDDVQVDPADPPIETATQAVIALPAR 360

QY 57 SEAAAVPVIGISQVRNMSKEK 80

DB 361 DKASGKSPQIDTSQEKRRKKKKX 384

RESULT 10

Q20429 Q20429 PRELIMINARY; PRT; 759 AA.

AC Q20429; 01-NOV-1996 (TrEMBLrel. 01, Created)

DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)

DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)

DE F45E12.2 protein.

GN F45E12.2.

OS Caenorhabditis elegans.

OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;

OC Rhabditidae; Peloderinae; Caenorhabditis.

OX NCBI\_TaxID=6239;

RN [1]

RP SEQUENCE FROM N.A.

RA STRAIN=Bristol N2;

RC MEDLINE=94150718; PubMed=7906398;

RA Wilson R., Ainscough R., Anderson K., Baynes C., Berke M., Coulson A.,

RA Bonfield J., Burton J., Connel M., Copsey T., Cooper J., Fulton L.,

RA Craxton M., Dear S., Du Z., Durbin R., Pavello A., Fulton L.,

RA Gardner A., Green P., Hawkins T., Hillier L., Jier M., Johnston L.,

RA Jones M., Kerhaw J., Kirsten J., Laister N., Latreille P.,

RA Lighthouse J., Lloyd C., McMurray A., Mortimore B., O'Callaghan M.,

RA Parsons J., Percy C., Rifkin L., Roopra A., Saunders D., Showkeen R.,

RA Smalton N., Smith A., Sonnhammer E., Staden K., Stalton J.,

RA Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K., Waterston R.,

RA Watson A., Weinstock L., Wilkinson-Sproat J., Wohlman P.,

RT \*2.2 Mb of contiguous nucleotide sequence from chromosome III of C.

```

RT RT Nature 368:32-38(1994).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RA Latreille P.;
RT "The sequence of C. elegans cosmid F45E12.";
RL Submitted (JUN-1995) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RA Waterston R.;
RL Submitted (JUN-1995) to the EMBL/GenBank/DBJ databases.
DR EMBL; U29536; AAA68790.1; -.
DR PIR; T16368; T16368.
DR WormPep; F45E12.2; CRO2738.
DR GO; GO:0005667; C:nucleus; IEA.
DR GO; GO:0003677; C:transcription factor complex; IEA.
DR GO; GO:0003702; F:RNA polymerase II transcription factor acti...; IEA.
DR GO; GO:0003355; P:regulation of transcription, DNA-dependent; IEA.
DR GO; GO:0006367; P:transcription initiation from Pol II promoter; IEA.
DR InterPro; IPR006670; Cyclin.
DR Pfam; PFO0382; transcript fac2; 2.
DR PRINTS; PRO0685; TIFACTOR1B.
DR SMART; SM00385; CYCLIN; 2.
SQ SEQUENCE 759 AA; 84033 MW; 16455477411D3043 CRC64;

Query Match
Best Local Similarity 30.4%; Score 87.5; DB 5; Length 759;
Matches 28; Conservative 17; Mismatches 32; Indels 15; Gaps 5;

QY 1 SGAGVPEKPCEDL--RCPTTSQALPAFTTEIQEASGPGADEV---QVPANAL 53
DB 577 ASETQKRSIDLTETSESKNSPKVNLKVESAS--PTSEVSSIEHPFVP---- 630

QY 54 PARSEAAVQPVIGISQVRMNSKEKDLGTL 85
DB 631 PAFSRVAKVPIIGAKXLAALN--EVKNVHTV 660

RESULT 11
Q9BL62 PRELIMINARY; PRT; 290 AA.
AC Q9BL62;
DT 01-JUN-2001 (T-EMBLrel. 17, Created)
DT 01-DEC-2001 (T-EMBLrel. 19, Last sequence update)
DT 01-OCT-2003 (T-EMBLrel. 25, Last annotation update)
DE Hypothetical protein.
GN Y65B4A.5.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RA Waterston R.;
RL Submitted (FEB-2003) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RA Maupin R.;
RL "Genome sequence of the nematode C. elegans: a platform for
investigating biology. The C. elegans Sequencing Consortium.";
RL Science 282:2012-2018(1998).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RA Wilson R.;
RL Submitted (JUN-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF067217; AAF99975.3; -.
KW Hypothetical protein.
SQ SEQUENCE 479 AA; 53887 MW; 7C56A024C7772C86 CRC64;

Query Match
Best Local Similarity 18.8%; Score 86; DB 5; Length 479;
Matches 30; Conservative 29; Mismatches 55; Indels 46; Gaps 4;

QY 7 PEKPCEDLRCPTTSQALPAFTTEIQEASGPGADEVQVFPANALPARSEAAVQFVI 66
DB 95 PEKR-----KLTDEKETQKHSHSENISYSPNSKFLKN-LKLIQLR 137

QY 67 GISQVRMNSKEKDLGTLGVLGVTMWIIAIGAGIILGYSYKRGKDLKEHQDKVCE 126
DB 138 ENAKRAFAEAEKERTL-----KLTDEKETQKHSHSENISYSPNSKFLKN-LKLIQLR 137

QY 127 REMQRTPLPSAFTNPCEIVDEKTVVHTSQTPVDPQEG 166
DB 173 KEIKRIEREISAATNSKCELY----TFCHVGKTVITWHG 208

RESULT 12
Q9GZG5 PRELIMINARY; PRT; 479 AA.
ID Q9GZG5;
AC Q9GZG5;
DT 01-MAR-2001 (T-EMBLrel. 15, Created)
DT 01-OCT-2003 (T-EMBLrel. 25, Last sequence update)
DT 01-OCT-2003 (T-EMBLrel. 25, Last annotation update)
DE Hypothetical protein F56A6.4.
GN F56A6.4.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RX MEDLINE=99069613; PubMed=9851916;
RA Wilson R.;
RT "Genome sequence of the nematode C. elegans: a platform for
investigating biology. The C. elegans Sequencing Consortium.";
RL Science 282:2012-2018(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RA Murray J.; Rohlfing T.; O'Neal D.; Wilson R.;
RT "The sequence of C. elegans cosmid F56A6.";
RL Submitted (MAR-1998) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RA Waterston R.;
RL Submitted (FEB-2003) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RA Wilson R.;
RL Submitted (JUN-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF067217; AAF99975.3; -.
KW Hypothetical protein.
SQ SEQUENCE 479 AA; 53887 MW; 7C56A024C7772C86 CRC64;

Query Match
Best Local Similarity 18.8%; Score 86; DB 5; Length 479;
Matches 30; Conservative 29; Mismatches 55; Indels 46; Gaps 4;

QY 7 PEKPCEDLRCPTTSQALPAFTTEIQEASGPGADEVQVFPANALPARSEAAVQFVI 66
DB 95 PEKR-----KLTDEKETQKHSHSENISYSPNSKFLKN-LKLIQLR 137

QY 67 GISQVRMNSKEKDLGTLGVLGVTMWIIAIGAGIILGYSYKRGKDLKEHQDKVCE 126
RT "Direct Submission.";
```

```

DE MKIAA1176 protein (Fragment).
GN Mus musculus (Mouse).
OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=22579291; PubMed=12693553;
RA Okazaki N., Kikuno T., Ohara R., Inamoto S., Aizawa H., Yuasa S.,
RA Nakajima D., Nagase T., Ohara O., Koga H.;
RT "Prediction of the coding sequences of mouse homologues of KIAA gene:
RT II. The complete nucleotide sequences of 400 mouse KIAA-homologous
RT cDNAs identified by screening of terminal sequences of cDNA clones
RT randomly sampled from size-fractionated libraries.";
RL DNA Res. 10:35-48(2003)
DR EMBL; AK122460; BAC65742.1; --
DR GO; GO:0016021; C:integral to membrane; IEA.
DR GO; GO:0005279; F:amino acid-polyamine transporter activity; IEA.
DR GO; GO:0015377; F:cation:chloride symporter activity; IEA.
DR GO; GO:0005215; F:transporter activity; IEA.
DR GO; GO:0006865; P:amino acid transport; IEA.
DR GO; GO:0006821; P:chloride transport; IEA.
DR GO; GO:0006811; P:ion transport; IEA.
DR GO; GO:0006814; P:sodium ion transport; IEA.
DR InterPro; IPR002293; AA/rel_permease1.
DR InterPro; IPR000076; KCL cotransp.
DR PRINTS; PR01081; KCLTRNSPORT.
DR TIGRFAMs; TIGR00930; 2a30; 1.
FT NON_TER 1
SQ SEQUENCE 1164 AA; 128729 MW; 20AA3B4F3D0B9F7 CRC64;

Query Match 9.3%; Score 85.5; DB 11; Length 1164;
Best Local Similarity 27.1%; Pred. No. 34;
Matches 32; Conservative 15; Mismatches 44; Indels 27; Gaps 4;

QY 79 KDLGTLGYVLGTMVVIITAGIGILGYSY-----KRGKDLKEQH-DQKVCERMQ 130
DB 871 KDLTTLFLHRLTAEEVVEVMEHSDISAYTEKTLVMEQSRQILKQHLTKNERREIQ 930
QY 131 RIT-----LPLSAFTNPTC---EIVDEKTVVHTSQTPVDQEGSTP 169
DB 931 SITDESRSIRRNPNRNLNVPETACDNEKPEEEVQLIHDSAPSCSPSPSP 1038

RESULT 15
QY Q8U3P9 PRELIMINARY; PRT; 607 AA.
AC Q8U3P9;
DT 01-JUN-2002 (TrEMBLrel. 21, Created)
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Hypothetical protein PF0407.
GN Pyrococcus furiosus.
OC Archaea; Euryarchaeota; Thermococci; Thermococcales; Thermococcaceae;
OX Pyrococcus.
OX NCBI_TaxID=2261;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Vcl / DSM 3638 / ATCC 43587 / JCM 8422;
RA Weiss R.B., Dunn D.M., Robb F.T., Brown J.R.;
RT "The complete sequence of the Pyrococcus furiosus genome.";
RL Submitted (FEB-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AE010163; AAL80531.1; --
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 607 AA; 68721 MW; 0D125AC09DE4A9AD CRC64;

Query Match 9.2%; Score 84.5; DB 17; Length 607;
Best Local Similarity 28.8%; Pred. No. 19;
Matches 36; Conservative 10; Mismatches 38; Indels 41; Gaps 6;

```

```

DB 138 ENAKGAREAEKRRKL-----EKDKKSEMARREKEK 172
QY 127 REMQRIITLPSAFTNPTCEIVDEKTVVHTSQTPVDQEG 166
DB 173 KETRKIERSAATNSKCELY---TFCHVGKVIDTWGH 208

RESULT 13
QY Q7TQC9 PRELIMINARY; PRT; 1114 AA.
AC Q7TQC9;
DT 01-OCT-2003 (TrEMBLrel. 25, Created)
DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Hypothetical protein.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6; TISSUE=Brain;
RX MEDLINE=22388257; PubMed=12477932;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Brownstein M.J., Ustin T.B., Toshiyuki S., Carninci P., Pange C.,
RA Rana S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullany S.J.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Kettelman M., Madan A., Rodriguez S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
RA Krzywinski M.I., Skalska J., Smutuz J., Schnerch A., Schein J.E.,
RA Jones S.J., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
RT and mouse cDNA sequences.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6; TISSUE=Brain;
RA Strausberg R.;
RL Submitted (JUL-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC054808; AAH54808.1; --
KW Hypothetical protein.
SQ SEQUENCE 1114 AA; 123471 MW; A8DE9C37F37CE444 CRC64;

Query Match 9.3%; Score 85.5; DB 11; Length 1114;
Best Local Similarity 27.1%; Pred. No. 33;
Matches 32; Conservative 15; Mismatches 44; Indels 27; Gaps 4;

QY 79 KDLGTLGYVLGTMVVIITAGIGILGYSY-----KRGKDLKEQH-DQKVCERMQ 130
DB 871 KDLTTLFLHRLTAEEVVEVMEHSDISAYTEKTLVMEQSRQILKQHLTKNERREIQ 930
QY 131 RIT-----LPLSAFTNPTC---EIVDEKTVVHTSQTPVDQEGSTP 169
DB 931 SITDESRSIRRNPNRNLNVPETACDNEKPEEEVQLIHDSAPSCSPSPSP 988

RESULT 14
QY Q80TIS PRELIMINARY; PRT; 1164 AA.
AC Q80TIS;
DT 01-JUN-2003 (TrEMBLrel. 24, Created)
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)

```



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 17, 2004, 06:55:13 ; Search time 45.7739 Seconds  
(without alignments)  
1104.909 Million cell updates/sec

Title: US-10-057-951-2\_COPY\_85\_263  
Perfect score: 916  
Sequence: 1 SGEAGVPEKPCDELRCPET.....PVDQEGSTPLMGQAGTGA 179

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : A Geneseq\_29Jan04:\*

- 1: Geneseqp1980s:\*
- 2: Geneseqp1980s:\*
- 3: Geneseqp2000s:\*
- 4: Geneseqp2001s:\*
- 5: Geneseqp2002s:\*
- 6: Geneseqp2003as:\*
- 7: Geneseqp2003bs:\*
- 8: Geneseqp2004s:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	916	100.0	263	2 AAW87769	Aaw87769 Human tis
2	916	100.0	263	2 AAY05219	Aay05219 Kringle1
3	916	100.0	263	4 AAE00300	Aae00300 Human tis
4	916	100.0	263	5 ABR40414	Abr40414 Human sec
5	913	99.7	263	3 AAB43237	Aab43237 Human ORF
6	913	99.7	263	5 AAU86149	Aau86149 Human PRO
7	902	98.5	263	4 AAM93748	Aam93748 Human pol
8	704.5	76.9	286	2 AAY05220	Aay05220 Kringle1
9	322	35.2	66	4 AAM18800	Aam18800 Peptide #
10	322	35.2	66	4 ABB37905	Abb37905 Peptide #
11	322	35.2	66	4 AAM31314	Aam31314 Peptide #
12	322	35.2	66	4 ABB23159	Abb23159 Protein #
13	322	35.2	66	4 AAM71037	Aam71037 Human bon
14	322	35.2	66	4 AAM58537	Aam58537 Human bra
15	322	35.2	66	4 ABG52752	Abg52752 Human liv
16	322	35.2	66	5 ABR40828	Abr40828 Human pep
17	275	30.0	146	5 ABR40487	Abr40487 Human sec
18	275	30.0	146	5 ABR40561	Abr40561 Human sec
19	95	10.4	163	4 AAU55844	Aau55844 Propionib
20	95	10.4	163	6 ABM52363	Abm52363 Propionib
21	94	10.3	8805	4 ABB67112	Abb67112 Drosophil
22	91	9.9	81	7 ABR42624	Abr42624 Human kri
23	86.5	9.4	507	2 AAY34756	Aay34756 C. pneumo
24	86.5	9.4	820	6 ABU26928	Abu26928 Protein e
25	86	9.4	2248	6 AAE37045	Aae37045 Human nuc

26	83	9.1	236	6 ABU00277	Abu00277 Human nov
27	80.5	8.8	214	5 ABUS1005	Abus1005 Helicobac
28	80.5	8.8	325	5 ABUS1212	Abus1212 Helicobac
29	80.5	8.8	416	5 ABUS1226	Abus1226 Helicobac
30	80.5	8.8	769	3 AAY53896	Aay53896 A Neisser
31	80	8.7	1237	6 ABUS5875	Abus5875 Human not
32	80	8.7	1237	6 AAE34032	Aae34032 Human not
33	80	8.7	1238	5 ABB07823	Abb07823 Human not
34	80	8.7	1238	6 ABB97801	Abb97801 Amino aci
35	80	8.7	1238	6 ABB72570	Abb72570 Human Not
36	80	8.7	1238	6 ABR61829	Abr61829 Human Jag
37	80	8.7	1238	7 ABR61758	Abr61758 Human Jag
38	80	8.7	1238	7 ADD01036	Add01036 Human Jag
39	80	8.7	1238	7 ADD01037	Add01037 Human Jag
40	80	8.7	1238	7 ADE57373	Ade57373 Human Pro
41	80	8.7	1238	7 ADE60834	Ade60834 Human Pro
42	79.5	8.7	769	3 AAY53897	Aay53897 A Neisser
43	79	8.6	1212	2 AAW44299	Aaw44299 Human ser
44	78.5	8.6	552	4 ABB71591	Abb71591 Drosophil
45	77.5	8.5	452	4 AAG82370	Aag82370 S. epider

ALIGNMENTS

RESULT 1

AAW87769  
ID AAW87769 standard; protein; 263 AA.

XX AC AAW87769;

XX DT 29-MAR-1999 (first entry)

XX DE Human tissue plasminogen activator-like protease t-PALP.

XX KW Tissue plasminogen activator-like protease; t-PALP; human;  
XX KW circulatory system-related disorder; blood clotting; stroke; thrombosis;  
XX KW peripheral arterial occlusion; pulmonary embolism; myocardiothrombosis;  
XX KW diagnosis; therapy.

XX OS Homo sapiens.

XX FH Key Location/Qualifiers

FT Peptide 1..21  
FT Protein /label= Sig\_peptide  
FT Peptide 22..263  
FT Protein /label= Mat\_protein  
FT Peptide 22..31  
FT Domain /note= "epitope-bearing region"  
FT Peptide 25..84  
FT Peptide /note= "kringle domain"  
FT Peptide 35..44  
FT Peptide /note= "epitope-bearing region"  
FT Peptide 71..81  
FT Domain /note= "epitope-bearing region"  
FT Peptide 85..263  
FT Peptide /note= "protease domain"  
FT Peptide 91..107  
FT Peptide /note= "epitope-bearing region"  
FT Peptide 119..128  
FT Peptide /note= "epitope-bearing region"  
FT Peptide 138..147  
FT Peptide /note= "epitope-bearing region"  
FT Peptide 155..167  
FT Peptide /note= "epitope-bearing region"  
FT Peptide 193..203  
FT Peptide /note= "epitope-bearing region"  
FT Peptide 206..215  
FT Peptide /note= "epitope-bearing region"  
FT Peptide 227..237  
FT Peptide /note= "epitope-bearing region"  
FT Peptide 243..252  
FT Peptide /note= "epitope-bearing region"

XX WO9854199-A1.  
 PN 03-DEC-1998.  
 PD 27-MAY-1998; 98WO-US010728.  
 XX 28-MAY-1997; 97US-0048000P.  
 PR (HUMA-) HUMAN GENOME SCI INC.  
 PA Ebner R, Moore PA, Ruben SM;  
 XX WPI; 1999-070207/06.  
 DR N-PSDB; AAV9636.  
 XX New tissue plasminogen activator-like protease - useful in the diagnosis  
 PT and treatment of circulatory system-related disorders.  
 XX Claim 1; Page 56-57; 76pp; English.  
 XX This is the amino acid sequence of tissue plasminogen activator-like  
 CC protease (t-PALP), a novel member of the serine protease family that  
 CC shares sequence homology to human tissue plasminogen activator (see  
 CC AAW87770). The t-PALP sequence was deduced from a cDNA clone (see  
 CC AAV9636) derived from activated monocytes. The 2.5 kb t-PALP message has  
 CC also been detected in heart, brain, lung, placenta, liver, skeletal  
 CC muscle, kidney, pancreas, spleen, thymus, prostate, testis, ovary, small  
 CC intestine, colon and peripheral blood leukocytes. Isolated nucleic acids  
 CC encoding amino acids -21 to 242, -20 to 242, 1-242, 4-63 (kringle domain)  
 CC and 64-242 (protease domain) of t-PALP, or encoding epitope-bearing  
 CC portions of t-PALP, are also claimed, as are recombinant vectors, host  
 CC cells, and methods for producing t-PALP polypeptides. t-PALP may be used  
 CC to detect and treat disorders related to the circulatory system, and to  
 CC identify agonists and antagonists of t-PALP activity. The homology  
 CC between t-PALP and tPA indicates that t-PALP may be involved in the  
 CC regulation of normal and abnormal clotting in e.g. stroke, deep-vein  
 CC thrombosis, peripheral arterial occlusion, pulmonary embolism and  
 CC myocardiothrombosis  
 XX Sequence 263 AA;  
 SQ  
 Query Match 100.0%; Score 916; DB 2; Length 263;  
 Best Local Similarity 100.0%; Pred. No. 1.1e-88;  
 Matches 179; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 SGEAGVPEKRPCEDLRCPTTSQALPAFTTEIQEASEGPGADEVQVFAPANALPARSEAA 60  
 Db 85 SGEAGVPEKRPCEDLRCPTTSQALPAFTTEIQEASEGPGADEVQVFAPANALPARSEAA 144  
 QY 61 AVQPVIGISQVRVNSKEKDLGTLGYVLGITMMVIIIAIGAGIILGYSKGDKLKEQH 120  
 Db 145 AVQPVIGISQVRVNSKEKDLGTLGYVLGITMMVIIIAIGAGIILGYSKGDKLKEQH 204  
 QY 121 DQKVCEREMQRIITPLSAFTNPTCEIVDEKTVVHTSQTPVDPQEGSTPLMGAGTPGA 179  
 Db 205 DQKVCEREMQRIITPLSAFTNPTCEIVDEKTVVHTSQTPVDPQEGSTPLMGAGTPGA 263  
 RESULT 2  
 AAY05219  
 ID AAY05219 standard; protein; 263 AA.  
 XX AAY05219;  
 AC AAY05219;  
 XX 17-JUN-1999 (first entry)  
 DT Kringle1 protein sequence.  
 DE Kringle1; cancer; inflammation; autoimmunity; rheumatoid arthritis; AIDS;  
 XX CNS inflammation; cerebellar degeneration; Alzheimer's disease; asthma;  
 KW Parkinson's disease; multiple sclerosis; amyotrophic lateral sclerosis;  
 KW neurological abnormality; ischaemia reperfusion injury; ischaemic injury;

KW cardiovascular disease; kidney disease; liver disease; aplastic anaemia;  
 KW myocardial infarction; hypotension; hypertension; allergy; infection;  
 KW myelodysplastic syndrome; haematologic abnormality; diagnosis; therapy;  
 KW male pattern baldness.  
 XX Homo sapiens.  
 OS WO9911788-A1.  
 PN 11-MAR-1999.  
 PD 02-SEP-1998; 98WO-US018270.  
 XX 02-SEP-1997; 97US-0056032P.  
 PR 01-SEP-1998; 98US-00144889.  
 XX (SMIK ) SMITHKLINE BEECHAM CORP.  
 PA Albone EF, Kikly KK;  
 XX WPI; 1999-214707/18.  
 DR N-PSDB; AAX28354.  
 XX New kringle1 polypeptides and polynucleotides.  
 PT Claim 1; Page 31-32; 42pp; English.  
 XX This sequence is a Kringle1 polypeptide of the invention. The kringle1  
 CC polypeptides (I) are used to screen for agonists and antagonists.  
 CC Agonists are used to treat subjects in need of enhanced activity or  
 CC expression of (I). Antagonists are used to treat subjects having need to  
 CC inhibit the activity or expression of (I). The methods can be used to  
 CC treat conditions such as cancer, inflammation, autoimmunity, allergy,  
 CC asthma, rheumatoid arthritis, CNS inflammation, cerebellar degeneration,  
 CC Alzheimer's disease, Parkinson's disease, multiple sclerosis,  
 CC amyotrophic lateral sclerosis, head injury damage and other neurological  
 CC abnormalities, ischaemia reperfusion injury, cardiovascular disease,  
 CC kidney disease, liver disease, ischaemic injury, myocardial infarction,  
 CC hypotension, hypertension, AIDS, myelodysplastic syndromes and other  
 CC haematologic abnormalities, aplastic anaemia, male pattern baldness, and  
 CC bacterial, fungal, protozoan and viral infections. The kringle1  
 CC polypeptides may also be used to generate antibodies. Determining the  
 CC presence or absence of mutations in, and analysing for the presence or  
 CC absence of expression of, kringle1 polynucleotides can be used to  
 CC diagnose a disease or susceptibility to a disease related to expression  
 CC or activity of kringle1 proteins. The polynucleotides may also be used  
 CC for chromosome identification, and mapping  
 XX Sequence 263 AA;  
 SQ  
 Query Match 100.0%; Score 916; DB 2; Length 263;  
 Best Local Similarity 100.0%; Pred. No. 1.1e-88;  
 Matches 179; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 SGEAGVPEKRPCEDLRCPTTSQALPAFTTEIQEASEGPGADEVQVFAPANALPARSEAA 60  
 Db 85 SGEAGVPEKRPCEDLRCPTTSQALPAFTTEIQEASEGPGADEVQVFAPANALPARSEAA 144  
 QY 61 AVQPVIGISQVRVNSKEKDLGTLGYVLGITMMVIIIAIGAGIILGYSKGDKLKEQH 120  
 Db 145 AVQPVIGISQVRVNSKEKDLGTLGYVLGITMMVIIIAIGAGIILGYSKGDKLKEQH 204  
 QY 121 DQKVCEREMQRIITPLSAFTNPTCEIVDEKTVVHTSQTPVDPQEGSTPLMGAGTPGA 179  
 Db 205 DQKVCEREMQRIITPLSAFTNPTCEIVDEKTVVHTSQTPVDPQEGSTPLMGAGTPGA 263  
 RESULT 3  
 AAE00300  
 ID AAE00300 standard; protein; 263 AA.  
 XX AAE00300;  
 AC AAE00300;  
 XX

13-JUN-2001 (first entry)  
 Human tissue-plasminogen activator-like protease (t-PALP).  
 Human: tissue-plasminogen activator-like protease; t-PALP; therapy;  
 vascular disease; stroke; deep vein thrombosis; keloid; asthma;  
 arterial occlusion; blood coagulation disorder; cerebroprotective;  
 autoimmune system disorder; human immunodeficiency syndrome; cytostatic;  
 rheumatoid arthritis; graft-versus-host disease; thyroiditis; cardiac;  
 insulin dependent diabetes; inflammatory eye disease; allergy; antiviral;  
 cardiovascular disease; heart disease; arrhythmia; myocardial ischaemia;  
 hyperproliferative disorder; hypertrophic scar; neurological disease;  
 Creutzfeldt-Jakob syndrome; neurodegenerative disorder; antiinflammatory;  
 Alzheimer's disease; Parkinson's disease; immunosuppressive; antifungal;  
 infectious disease; drug screening; gene therapy; neuroprotective;  
 cancer; ophthalmological; antibacterial; vulnery.  
 XX  
 OS Homo sapiens.  
 XX  
 XX  
 FH Key Location/Qualifiers  
 FT Binding-site 1..165  
 FT /note= "Binds to FLAG polypeptide to form t-PALP-FLAG  
 FT fusion protein"  
 FT Peptide 1..21  
 FT /label= Signal\_peptide  
 FT Domain 4..63  
 FT /label= Kringle\_domain  
 FT Region 12..21  
 FT /note= "Conserved region"  
 FT Protein 22..263  
 FT /note= "Human mature tissue-plasminogen activator-like  
 FT protease (t-PALP); Binds to FLAG polypeptide to form t-  
 FT PALP-FLAG fusion protein"  
 FT Region 22..38  
 FT /note= "Conserved region"  
 FT Region 22..31  
 FT /note= "Epitope-bearing portion"  
 FT Region 35..44  
 FT /note= "Epitope-bearing portion"  
 FT Region 39..49  
 FT /note= "Conserved region"  
 FT Region 50..62  
 FT /note= "Conserved region"  
 FT Region 63..84  
 FT /note= "Conserved region"  
 FT Domain 64..242  
 FT /label= Protease\_domain  
 FT Region 71..81  
 FT /note= "Epitope-bearing portion"  
 FT Region 85..97  
 FT /note= "Conserved region"  
 FT Region 91..107  
 FT /note= "Epitope-bearing portion"  
 FT Region 100..118  
 FT /note= "Conserved region"  
 FT Region 119..128  
 FT /note= "Epitope-bearing portion"  
 FT Region 119..127  
 FT /note= "Conserved region"  
 FT Region 128..143  
 FT /note= "Conserved region"  
 FT Region 138..147  
 FT /note= "Epitope-bearing portion"  
 FT Region 146..163  
 FT /note= "Conserved region"  
 FT Region 155..167  
 FT /note= "Epitope-bearing portion"  
 FT Region 164..180  
 FT /note= "Conserved region"  
 FT Region 186..200  
 FT /note= "Conserved region"  
 FT Region 193..203  
 FT /note= "Epitope-bearing portion"

FT Region 201..220  
 FT /note= "Conserved region"  
 FT Region 206..215  
 FT /note= "Epitope-bearing portion"  
 FT Region 221..236  
 FT /note= "Conserved region"  
 FT Region 227..237  
 FT /note= "Epitope-bearing portion"  
 FT Region 237..248  
 FT /note= "Conserved region"  
 FT Region 243..252  
 FT /note= "Epitope-bearing portion"  
 FT Region 249..263  
 FT /note= "Conserved region"  
 XX WO200125252-A1.  
 PN 12-APR-2001.  
 XX  
 PD 03-OCT-2000; 2000WO-US027239.  
 XX  
 PF 04-OCT-1999; 99US-00411977.  
 PR  
 XX (HUMA-) HUMAN GENOME SCI INC.  
 PA  
 XX Moore PA, Ruben SM, Ebner R;  
 PI WPI, 2001-235402/24.  
 XX DR N-PSDB; AAD03460.  
 XX  
 PT New (gene encoding and antibody immunospecific for a) tissue-plasminogen  
 PT activator-like protease, useful for the diagnosis and treatment of  
 PT (cardio)vascular diseases, hyperproliferative disorders, immune system  
 PT disorders and cancers.  
 PT  
 XX Claim 17; Fig 1; 323pp; English.  
 PS  
 XX The present amino acid sequence is HMTS42 clone human tissue-plasminogen  
 CC activator-like protease (t-PALP). The t-PALP sequence and their  
 CC (ant)agonists are useful for the diagnosis and treatment of vascular  
 CC diseases e.g. stroke, deep vein thrombosis and arterial occlusion, blood  
 CC coagulation disorders, (auto)immune system disorders e.g. human  
 CC immunodeficiency syndrome, rheumatoid arthritis, graft-versus-host  
 CC disease, thyroiditis, insulin dependent diabetes and inflammatory eye  
 CC disease, allergic reactions e.g. asthma, cardiovascular diseases e.g.  
 CC heart disease, arrhythmia and myocardial ischaemia, hyperproliferative  
 CC disorders, cancers, hypertrophic scars and keloids, neurological diseases  
 CC e.g. Creutzfeldt-Jakob syndrome, neurodegenerative disorders e.g.  
 CC Alzheimer's disease and Parkinson's disease and infectious disease e.g.  
 CC viral, bacterial and fungal infections. The t-PALP sequences are also  
 CC useful for drug screening. The t-PALP nucleotides are useful as  
 CC chromosome markers and are involved in gene therapy  
 XX  
 SQ Sequence 263 AA;  
 Query Match 100.0%; Score 915; DB 4; Length 263;  
 Best Local Similarity 100.0%; Pred. No. 1.1e-88;  
 Matches 179; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 SGEAGVPEKPCEDLRCPETTSQALPAFTTEIOEASEGPGADEVQVFAPANALPARSEAA 60  
 DB 85 SGEAGVPEKPCEDLRCPETTSQALPAFTTEIOEASEGPGADEVQVFAPANALPARSEAA 144  
 QY 61 AVQPVIGISQVRMNSKEKDLGLTGLVGLITMMVIIIAIGAGIILGYSYKRGKDLKEQH 120  
 DB 145 AVQPVIGISQVRMNSKEKDLGLTGLVGLITMMVIIIAIGAGIILGYSYKRGKDLKEQH 204  
 QY 121 DQKVCEREMORITLPLSAFTNPTCEIVDEKTVVHTSOTPTVPOEASTPLMGAGTPGA 179  
 DB 205 DQKVCEREMORITLPLSAFTNPTCEIVDEKTVVHTSOTPTVPOEASTPLMGAGTPGA 263  
 RESULT 4





CC sequences have activities such as: cytostatic; hepatotropic; vulnery; CC antipariatic; antiparkinsonian; nootropic; neuroprotective; osteopathic; CC anticonvulsant; antiarthritic; immunosuppressant; immunostimulant; CC cardiact; thrombolytic; coagulant; vasotropic; antidiabetic; hypotensive; CC dermatological; immunosuppressive; antinflammatory; antibacterial; CC antiviral; antifungal; antineumatic; antithyroid; and antianemic. The CC sequences can be used for determining the presence of or predisposition to, or preventing or treating pathological conditions associated with an ORFX-associated disorder. The nucleic acids can be used to express ORFX proteins in gene therapy vectors. The proteins and nucleic acids may be used to treat cancers, proliferative disorders, neurodegenerative disorders, osteoarthritis, graft vs host disease, cardiovascular disease, diabetes mellitus, hypertension, hypothyroidism, cholesterol ester storage, systemic lupus erythematosus, severe combined immunodeficiency (SCID), AIDS, viral, bacterial or fungal infection, malaria, autoimmune disorders, asthma, allergies, aplastic anaemia, burns, wounds, bone and cartilage damage, nocturnal haemoglobinuria, antinflammatory disease; to enhance coagulation; to inhibit thrombosis; and as a contraceptive

XX SQ Sequence 263 AA;

Query Match 99.7%; Score 913; DB 3; Length 263;  
Best Local Similarity 99.4%; Pred. No. 2.3e-88;  
Matches 178; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
QY 1 SGEAGVPEKPCEDLRCPETTSQALPAFTTEIQEASGPGADVQVPAPANALPARSEAA 60  
DB 85 SGEAGVPEKPCEDLRCPETTSQALPAFTTEIQEASGPGADVQVPAPANALPARSEAA 144  
QY 61 AVQPVIGISQVRNRSKEXDGLTGLVGLGTTMMVIIAIGAGIILGYSYKRGKDLKEQH 120  
DB 145 AVQPVIGISQVRNRSKEXDGLTGLVGLGTTMMVIIAIGAGIILGYSYKRGKDLKEQH 204  
QY 121 DQKVCEREMQRTITPLSAFTNPTCEIVDEKTVVHTSQTVPDQEGSTPLMGQAGTPGA 179  
DB 205 DQKVCEREMQRTITPLSAFTNPTCEIVDEKTVVHTSQTVPDQEGSTPLMGQAGTPGA 263

RESULT 6  
AAU86149  
ID AAU86149 standard; protein; 263 AA.  
AC AAU86149;  
XX  
DT 15-JUL-2002 (first entry)  
XX  
DE Human PRO264 polypeptide.  
XX  
KW Human; PRO; benign tumour; malignant tumour; lymphoid malignancy;  
KW leukaemia; neuronal disorder; stromal disorder; blastocoeic disorder;  
KW inflammatory disorder; immune disorder; angiogenic disorder; cytostatic;  
KW neuroprotective.  
XX  
OS Homo sapiens.  
XX  
PN WO200153486-A1.  
XX  
PD 26-JUL-2001.  
XX  
PF 11-FEB-2000; 2000WO-US003565.  
XX  
PR 08-MAR-1999; 99WO-US005028.  
PR 11-MAR-1999; 99US-0123972P.  
PR 11-MAY-1999; 99US-0133459P.  
PR 02-JUN-1999; 99WO-US012252.  
PR 22-JUN-1999; 99US-0140650P.  
PR 20-JUL-1999; 99US-0144758P.  
PR 26-JUL-1999; 99US-0145698P.  
PR 28-JUL-1999; 99US-0146222P.  
PR 17-AUG-1999; 99US-0149395P.  
PR 31-AUG-1999; 99US-0151689P.  
PR 01-SEP-1999; 99WO-US020111.

PR 15-SEP-1999; 99WO-US021090.  
PR 30-NOV-1999; 99WO-US028313.  
PR 01-DEC-1999; 99WO-US028301.  
PR 01-DEC-1999; 99WO-US028634.  
PR 05-JAN-2000; 2000WO-US000219.  
XX (GETH ) GENENTECH INC.  
PA  
XX Ashkenazi AJ, Goddard A, Godowski PJ, Gurney AL, Hillan KJ;  
PI Marsters SA, Pan J, Pitti RM, Roy MA, Smith V, Stone DM;  
PI Watanabe CK, Wood WI;  
XX  
DR WPI: 2002-205567/26.  
DR N-PSDB; ABK40275.  
XX  
PT Thirty five nucleic acids encoding PRO polypeptides, useful for treating  
PT benign or malignant tumors, leukemias and lymphoid malignancies,  
PT inflammatory, angiogenic and immunologic disorders.  
XX  
PS Claim 61; Fig 44; 302pp; English.  
XX  
CC The present invention relates to the isolation of novel human PRO  
CC polypeptides and the polynucleotide sequences encoding them. The PRO  
CC polypeptides, agonists, antagonists or anti-PRO antibodies are useful for  
CC treating benign or malignant tumours (e.g. renal, kidney, bladder,  
CC breast etc) leukaemias and lymphoid malignancies, other disorders such  
CC as neuronal, glial, astrocytal, hypothalamic, glandular, macrophagal,  
CC stromal and blastocoeic disorders, inflammatory, immune and angiogenic  
CC disorders. The polynucleotide sequences are also useful in gene therapy.  
CC AAU86128-AAU86162 represent the human PRO polypeptides of the invention  
XX  
SQ Sequence 263 AA;

Query Match 99.7%; Score 913; DB 5; Length 263;  
Best Local Similarity 99.4%; Pred. No. 2.3e-88;  
Matches 178; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
QY 1 SGEAGVPEKPCEDLRCPETTSQALPAFTTEIQEASGPGADVQVPAPANALPARSEAA 60  
DB 85 SGEAGVPEKPCEDLRCPETTSQALPAFTTEIQEASGPGADVQVPAPANALPARSEAA 144  
QY 61 AVQPVIGISQVRNRSKEXDGLTGLVGLGTTMMVIIAIGAGIILGYSYKRGKDLKEQH 120  
DB 145 AVQPVIGISQVRNRSKEXDGLTGLVGLGTTMMVIIAIGAGIILGYSYKRGKDLKEQH 204  
QY 121 DQKVCEREMQRTITPLSAFTNPTCEIVDEKTVVHTSQTVPDQEGSTPLMGQAGTPGA 179  
DB 205 DQKVCEREMQRTITPLSAFTNPTCEIVDEKTVVHTSQTVPDQEGSTPLMGQAGTPGA 263  
RESULT 7  
AAU93748  
ID AAU93748 standard; protein; 263 AA.  
XX  
AC AAU93748;  
XX  
DT 06-NOV-2001 (first entry)  
XX  
DE Human polypeptide, SEQ ID NO: 3727.  
XX  
KW Human; full length cDNA; cDNA synthesis; oligo-capping.  
XX  
OS Homo sapiens.  
XX  
PN EP1130094-A2.  
XX  
PD 05-SEP-2001.  
XX  
PF 07-JUL-2000; 2000EP-00114089.  
XX  
PR 08-JUL-1999; 99JP-00194486.  
PR 11-JAN-2000; 2000JP-00118774.  
PR 02-MAY-2000; 2000JP-00183765.

```

XX PA (HELI-) HELIX RES INST.
XX PI Ota T, Nishikawa T, Isogai T, Hayaashi K, Ishii S, Kawai Y;
XX PI Wakamatsu A, Sugiyama T, Nagai K, Kojima S, Otsuki T, Koga H;
XX DR WPI; 2001-524255/58.
XX DR N-PSDB; AAK94700.
XX PT 830 Primers useful for synthesizing full length cDNA clones and their use
XX PT in genetic manipulation.
XX PS Claim 8; SEQ ID NO 3727; 1380pp + Sequence Listing; English.
XX CC The invention relates to primers for synthesising full length cDNA
XX CC clones. 830 cDNA molecules encoding a human protein have been isolated
XX CC and nucleotide sequences of 5' - and 3' - ends of the cDNA molecules have
XX CC been determined. Primers for synthesising the full length cDNA are useful
XX CC for clarifying the function of the protein encoded by the cDNA. The full
XX CC length clones were obtained by construction of full length enriched cDNA
XX CC libraries that were synthesised by the oligo-capping method. The primers
XX CC enable the production of the full length cDNA easily without any special
XX CC methods. The present sequence is a polypeptide encoded by a full length
XX CC human cDNA of the invention. Note: The sequence data for this patent did
XX CC not form part of the printed specification, but was obtained in CD-ROM
XX CC format directly from EPO
XX SQ Sequence 263 AA;
Query Match 98.5%; Score 902; DB 4; Length 263;
Best Local Similarity 98.9%; Pred. No. 3.4e-87;
Matches 177; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 1 SGEAGVPEKRCEDLRCPTTSQALPAFTTEIQEASEGPGADEVQVFPANALPARSEAA 60
DB 85 SGEAGVPEKRCEDLRCPTTSQALPAFTTEIQEASEGPGADEVQVFPANALPARSEAA 144
QY 61 AVQPVIGISQVRNNSKEKDLGLTGYVLGITMWWIIIAICAGIILGYSYKRGKDLKEQH 120
DB 145 AVQPVIGISQVRNNSKEKDLGLTGYVLGITMWWIIIAICAGIILGYSYKRGKDLKEQH 204
QY 121 DQKVCEREMQRIITPLSAFTNPTCEIVDEKTVVHTSQTVPDPOEGSTPLMGAGTPGA 179
DB 205 DQKVCEREMQRIITPLSAFTNPTCEIVDEKTVVHTSQTVPDPOEGSTPLMGAGTPGA 263
RESULT 8
RAY05220
ID AAY05220 standard; protein; 286 AA.
XX AC AAY05220;
XX DT 17-JUN-1999 (first entry)
XX DE Kringlei protein sequence.
XX KW Kringlei; cancer; inflammation; autoimmunity; rheumatoid arthritis; AIDS;
XX KW CNS inflammation; cerebellar degeneration; Alzheimer's disease; asthma;
XX KW Parkinson's disease; multiple sclerosis; amyotrophic lateral sclerosis;
XX KW neurological abnormality; ischaemia reperfusion injury; ischaemic injury;
XX KW cardiovascular disease; kidney disease; liver disease; aplastic anaemia;
XX KW myocardial infarction; hypotension; hypertension; allergy; infection;
XX KW myelodysplastic syndrome; haematologic abnormality; diagnosis; therapy;
XX KW male pattern baldness.
XX OS Homo sapiens.
XX FN WO9911788-A1.
XX PD 11-MAR-1999.
XX PF 02-SEP-1998; 98WO-US018270.
XX

```

```

PR 02-SEP-1997; 97US-0056032P.
PR 01-SEP-1998; 98US-00144889.
XX PA (SMK ) SMITHKLINE BEECHAM CORP.
XX PI Albone EF, Kikly KK;
XX WPI; 1999-214707/18.
XX DR N-PSDB; AAX28355.
XX PT New kringlei polypeptides and polynucleotides.
XX PS Claim 14; Page 33; 42pp; English.
XX CC This sequence is a Kringlei polypeptide of the invention. The kringlei
XX CC polypeptides (I) are used to screen for agonists and antagonists.
XX CC Agonists are used to treat subjects in need of enhanced activity or
XX CC expression of (I). Antagonists are used to treat subjects having need
XX CC to inhibit the activity or expression of (I). The methods can be used to
XX CC treat conditions such as cancer, inflammation, autoimmunity, allergy,
XX CC asthma, rheumatoid arthritis, CNS inflammation, cerebellar degeneration,
XX CC Alzheimer's disease, Parkinson's disease, multiple sclerosis,
XX CC amyotrophic lateral sclerosis, head injury damage and other neurological
XX CC abnormalities, ischaemia reperfusion injury, cardiovascular disease,
XX CC kidney disease, liver disease, ischaemic injury, myocardial infarction,
XX CC hypotension, hypertension, AIDS, myelodysplastic syndromes and other
XX CC haematologic abnormalities, aplastic anaemia, male pattern baldness, and
XX CC bacterial, fungal, protozoan and viral infections. The kringlei
XX CC polypeptides may also be used to generate antibodies. Determining the
XX CC presence or absence of mutations in, and analysing for the presence or
XX CC absence of expression of, kringlei polynucleotides can be used to
XX CC diagnose a disease or susceptibility to a disease related to expression
XX CC or activity of kringlei proteins. The polynucleotides may also be used
XX CC for chromosome identification, and mapping
XX SQ Sequence 286 AA;
Query Match 76.9%; Score 704.5; DB 2; Length 286;
Best Local Similarity 82.4%; Pred. No. 3.7e-66;
Matches 145; Conservative 5; Mismatches 25; Indels 1; Gaps 1;
QY 1 SGEAGVPEKRCEDLRCPTTSQALPAFTTEIQEASEGPGADEVQVFPANALPARSEAA 59
DB 85 SGEAGVPEKRCEDLRCPTTSQALPAFTTEIQEASEGPGADEVQVFPANALPARSEAA 144
QY 60 AAVQPVIGISQVRNNSKEKDLGLTGYVLGITMWWIIIAICAGIILGYSYKRGKDLKEQ 119
DB 145 ALLOPVIGISQVRDELOGEKGPGNSGLRACHYHGDIIAIGAGIILGYSYKRGKDLKEQ 204
QY 120 HDQKVCEREMQRIITPLSAFTNPTCEIVDEKTVVHTSQTVPDPOEGSTPLMGAG 175
DB 205 HDQKVCEREMQRIITPLSAFTNPTCEIVDEKTVVHTSQTVPDPOEGSTPLMGAG 260
RESULT 9
AAM18800
ID AAM18800 standard; protein; 66 AA.
XX AC AAM18800;
XX DT 12-OCT-2001 (first entry)
XX DE Peptide #5234 encoded by probe for measuring cervical gene expression.
XX KW Probe; human; microarray; gene expression; cervical epithelial cell;
XX KW cervical cancer.
XX OS Homo sapiens.
XX FN WO2000157278-A2.
XX PD 09-AUG-2001.
XX

```

04-OCT-2000; 2000GB-00024263.  
(MOLE-) MOLECULAR DYNAMICS INC.  
Penn SG, Hanzel DK, Chen W, Rank DR;  
WPI; 2001-483447/52.  
Human genome-derived single exon nucleic acid probes useful for analyzing gene expression in human fetal liver.  
Claim 27; SEQ ID NO 30540; 639pp + Sequence Listing; English.  
The invention relates to a single exon nucleic acid probe for measuring human gene expression in a sample derived from human foetal liver. The single exon nucleic acid probes may be used for predicting, measuring and displaying gene expression in samples derived from human fetal liver. The present sequence is a peptide encoded by a single exon nucleic acid probe of the invention. Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published\_pct\_sequences

Query Match 35.2%; Score 322; DB 4; Length 66;  
Best Local Similarity 100.0%; Pred. No. 2.2e-26;  
Matches 66; Conservative 0; Mismatches 0; Indels 0; Gaps 0

20 TTSQLPAPTFTTIOEASGPGADEVQVFPANALPARSEAAAVQVIGISQVRMNSKEK 79  
1 TTSQLPAPTFTTIOEASGPGADEVQVFPANALPARSEAAAVQVIGISQVRMNSKEK 60

80 KDLGTL 85  
61 KDLGTL 66

RESULT 11  
AAM31314  
ID AAM31314 standard; protein; 66 AA.  
AC AAM31314;  
XX  
DT 17-OCT-2001 (first entry)  
DE  
XX Peptide #5351 encoded by probe for measuring placental gene expression.  
XX Probe; microarray; human; placenta; antenatal diagnosis;  
XX genetic disorder.  
XX Homo sapiens.  
XX WO200157272-A2.  
XX  
XX 09-AUG-2001.  
XX  
XX 30-JAN-2001; 2001WO-US000663.  
XX  
XX 04-FEB-2000; 2000US-0180312P.  
XX 26-MAY-2000; 2000US-0207456P.  
XX 30-JUN-2000; 2000US-00608408.  
XX 03-AUG-2000; 2000US-00632366.  
XX 21-SEP-2000; 2000US-0234687P.  
XX 27-SEP-2000; 2000US-0236359P.  
XX 04-OCT-2000; 2000GB-00024263.  
(MOLE-) MOLECULAR DYNAMICS INC.  
Penn SG, Hanzel DK, Chen W, Rank DR;  
WPI; 2001-488997/53.  
Human genome-derived single exon nucleic acid probes useful for analyzing gene expression in human foetal liver single exon probe.

Query Match 35.2%; Score 322; DB 4; Length 66;  
Best Local Similarity 100.0%; Pred. No. 2.2e-26;  
Matches 66; Conservative 0; Mismatches 0; Indels 0; Gaps 0

20 TTSQLPAPTFTTIOEASGPGADEVQVFPANALPARSEAAAVQVIGISQVRMNSKEK 79  
1 TTSQLPAPTFTTIOEASGPGADEVQVFPANALPARSEAAAVQVIGISQVRMNSKEK 60

80 KDLGTL 85  
61 KDLGTL 66

RESULT 10  
ABB37905  
ID ABB37905 standard; peptide; 66 AA.  
XX  
XX ABB37905;  
XX  
XX 04-FEB-2000 (first entry)  
XX  
XX Peptide #5411 encoded by human foetal liver single exon probe.  
XX Human; foetal liver; gene expression; single exon nucleic acid probe.  
XX Homo sapiens.  
XX  
XX WO200157277-A2.  
XX  
XX 09-AUG-2001.  
XX  
XX 30-JAN-2001; 2001WO-US000669.  
XX  
XX 04-FEB-2000; 2000US-0180312P.  
XX 26-MAY-2000; 2000US-0207456P.  
XX 30-JUN-2000; 2000US-00608408.  
XX 03-AUG-2000; 2000US-00632366.  
XX 21-SEP-2000; 2000US-0234687P.  
XX 27-SEP-2000; 2000US-0236359P.

PT gene expression in human placenta.  
PS Claim 27; SEQ ID NO 31593; 654pp; English.  
XX  
CC The present invention relates to single exon nucleic acid probes (SENPs;  
CC see AA131315-AA157546). The present sequence is a peptide encoded by one  
CC such probe. The probes are useful for producing a microarray for  
CC predicting, measuring and displaying gene expression in samples derived  
CC from human placenta. The probes are useful for antenatal diagnosis of  
CC human genetic disorders  
XX  
SQ Sequence 66 AA;  
Query Match 35.2%; Score 322; DB 4; Length 66;  
Best Local Similarity 100.0%; Pred. No. 2.2e-26;  
Matches 66; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 20 TTSQLPAPFTTEIQASEGPGADEVQVFAPANALPARSEAAAVQVIGISQVRMNSKEK 79  
Db 1 TTSQLPAPFTTEIQASEGPGADEVQVFAPANALPARSEAAAVQVIGISQVRMNSKEK 60  
QY 80 KDLGTL 85  
Db 61 KDLGTL 66  
RESULT 12  
ASB23159  
ID ABB23159 standard; protein; 66 AA.  
XX  
AC ABB23159;  
XX  
DT 23-JAN-2002 (first entry)  
XX  
DE Protein #5158 encoded by probe for measuring heart cell gene expression.  
XX  
KW Human; gene expression; heart; microarray; vascular system;  
KW cardiovascular disease; hypertension; cardiac arrhythmia;  
KW congenital heart disease.  
XX  
OS Homo sapiens.  
XX  
PN WO200157274-A2.  
XX  
PD 09-AUG-2001.  
XX  
PF 30-JAN-2001; 2001WO-US000666.  
XX  
PR 04-FEB-2000; 2000US-0180312P.  
PR 26-MAY-2000; 2000US-0207456P.  
PR 30-JUN-2000; 2000US-00608408.  
PR 03-AUG-2000; 2000US-00632366.  
PR 21-SEP-2000; 2000US-0234687P.  
PR 27-SEP-2000; 2000US-0236359P.  
PR 04-OCT-2000; 2000GB-00024263.  
XX  
PA (MOLE-) MOLECULAR DYNAMICS INC.  
XX  
PI Penn SG, Hanzel DK, Chen W, Rank DR;  
XX  
WPI; 2001-48899/53.  
XX  
PT Single exon nucleic acid probes for analyzing gene expression in human  
PT hearts.  
XX  
PS Claim 15; SEQ ID NO 24929; 530pp; English.  
XX  
CC The present invention relates to single exon nucleic acid probes for  
CC measuring human gene expression in a sample derived from human heart (see  
CC ABA21535-ABA41305). The present sequence is a protein encoded by one such  
CC probe. The probes may be used for predicting, measuring and displaying  
CC gene expression in samples derived from the human heart via microarrays.  
CC By measuring gene expression, the probes are useful for predicting,

CC diagnosing, grading, staging, monitoring and prognosing diseases of the  
CC human heart and vascular system e.g. cardiovascular disease,  
CC hypertension, cardiac arrhythmias and congenital heart disease. Note: The  
CC sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences  
XX  
SQ Sequence 66 AA;  
Query Match 35.2%; Score 322; DB 4; Length 66;  
Best Local Similarity 100.0%; Pred. No. 2.2e-26;  
Matches 66; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 20 TTSQLPAPFTTEIQASEGPGADEVQVFAPANALPARSEAAAVQVIGISQVRMNSKEK 79  
Db 1 TTSQLPAPFTTEIQASEGPGADEVQVFAPANALPARSEAAAVQVIGISQVRMNSKEK 60  
QY 80 KDLGTL 85  
Db 61 KDLGTL 66  
RESULT 13  
AMW71037  
ID AMW71037 standard; protein; 66 AA.  
XX  
AC AMW71037;  
XX  
DT 06-NOV-2001 (first entry)  
XX  
DE Human bone marrow expressed probe encoded protein SEQ ID NO: 31343.  
XX  
KW Human; bone marrow expressed exon; gene expression analysis; probe;  
KW microarray; cancer; leukaemia; lymphoma; myeloma.  
XX  
OS Homo sapiens.  
XX  
PN WO200157276-A2.  
XX  
PD 09-AUG-2001.  
XX  
PF 30-JAN-2001; 2001WO-US000668.  
XX  
PR 04-FEB-2000; 2000US-0180312P.  
PR 26-MAY-2000; 2000US-0207456P.  
PR 30-JUN-2000; 2000US-00608408.  
PR 03-AUG-2000; 2000US-00632366.  
PR 21-SEP-2000; 2000US-0234687P.  
PR 27-SEP-2000; 2000US-0236359P.  
PR 04-OCT-2000; 2000GB-00024263.  
XX  
PA (MOLE-) MOLECULAR DYNAMICS INC.  
XX  
PI Penn SG, Hanzel DK, Chen W, Rank DR;  
XX  
WPI; 2001-48899/53.  
XX  
PT Human genome-derived single exon nucleic acid probes useful for analyzing  
PT gene expression in human bone marrow.  
XX  
PS Example 4; SEQ ID NO 31343; 658pp + Sequence Listing; English.  
XX  
CC The present invention provides a number of single exon nucleic acid  
CC probes which are derived from genomic sequences expressed in the human  
CC bone marrow. They can be used to measure gene expression in bone marrow  
CC samples, which may enable the improved diagnosis and treatment of cancers  
CC such as lymphoma, leukaemia and myeloma. The present sequence is a  
CC protein encoded by one of the probes of the invention.  
XX  
SQ Sequence 66 AA;  
Query Match 35.2%; Score 322; DB 4; Length 66;  
Best Local Similarity 100.0%; Pred. No. 2.2e-26;

Matches 66; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 20 TTSQLPAFTTETIQEASEGPGADEVQVFAPANALPARSEAAAVQPVIGISQVRMNSKEK 79  
ID AAM58537 standard; protein; 66 AA.  
XX  
AC AAM58537;  
XX  
DT 05-NOV-2001 (first entry)  
XX  
DE Human brain expressed single exon probe encoded protein SEQ ID NO: 30642.  
XX  
KW Human; brain expressed exon; gene expression analysis; probe; microarray;  
KW Alzheimer's disease; multiple sclerosis; schizophrenia; epilepsy; cancer.  
XX  
OS Homo sapiens.  
XX  
PN WO200157275-A2.  
XX  
PD 09-AUG-2001.  
XX  
PF 30-JAN-2001; 2001WO-US000667.  
XX  
PR 04-FEB-2000; 2000US-0180312P.  
XX  
PR 26-MAY-2000; 2000US-0207456P.  
XX  
PR 30-JUN-2000; 2000US-00608408.  
XX  
PR 03-AUG-2000; 2000US-00632366.  
XX  
PR 21-SEP-2000; 2000US-0234687P.  
XX  
PR 27-SEP-2000; 2000US-0236359P.  
XX  
PR 04-OCT-2000; 2000GB-00024263.  
XX  
PA (MOLE-) MOLECULAR DYNAMICS INC.  
XX  
PI Penn SG, Hanzel DK, Chen W, Rank DR;  
XX  
XX WPI; 2001-483446/52.  
XX  
XX Single exon nucleic acid probes for analyzing gene expression in human  
XX brains.  
XX  
XX Example 4; SEQ ID NO 30642; 650pp + Sequence Listing; English.  
XX  
XX The present invention provides a number of single exon nucleic acid  
XX probes which are derived from genomic sequences expressed in the human  
XX brain. They can be used to measure gene expression in brain cell samples,  
XX which may enable the diagnosis and improved treatment of nervous system  
XX diseases such as Alzheimer's disease, multiple sclerosis, schizophrenia,  
XX epilepsy and cancers. The present sequence is a protein encoded by one of  
XX the probes of the invention

Query Match 35.2%; Score 322; DB 4; Length 66;  
Best Local Similarity 100.0%; Pred. No. 2.2e-26;  
Matches 66; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 20 TTSQLPAFTTETIQEASEGPGADEVQVFAPANALPARSEAAAVQPVIGISQVRMNSKEK 79  
ID AAM58537 standard; protein; 66 AA.  
XX  
AC AAM58537;  
XX  
DT 05-NOV-2001 (first entry)  
XX  
DE Human brain expressed single exon probe encoded protein SEQ ID NO: 30642.  
XX  
KW Human; brain expressed exon; gene expression analysis; probe; microarray;  
KW Alzheimer's disease; multiple sclerosis; schizophrenia; epilepsy; cancer.  
XX  
OS Homo sapiens.  
XX  
PN WO200157275-A2.  
XX  
PD 09-AUG-2001.  
XX  
PF 30-JAN-2001; 2001WO-US000667.  
XX  
PR 04-FEB-2000; 2000US-0180312P.  
XX  
PR 26-MAY-2000; 2000US-0207456P.  
XX  
PR 30-JUN-2000; 2000US-00608408.  
XX  
PR 03-AUG-2000; 2000US-00632366.  
XX  
PR 21-SEP-2000; 2000US-0234687P.  
XX  
PR 27-SEP-2000; 2000US-0236359P.  
XX  
PR 04-OCT-2000; 2000GB-00024263.  
XX  
PA (MOLE-) MOLECULAR DYNAMICS INC.  
XX  
PI Penn SG, Hanzel DK, Chen W, Rank DR;  
XX  
XX WPI; 2001-483446/52.  
XX  
XX Single exon nucleic acid probes for analyzing gene expression in human  
XX brains.  
XX  
XX Example 4; SEQ ID NO 30642; 650pp + Sequence Listing; English.  
XX  
XX The present invention provides a number of single exon nucleic acid  
XX probes which are derived from genomic sequences expressed in the human  
XX brain. They can be used to measure gene expression in brain cell samples,  
XX which may enable the diagnosis and improved treatment of nervous system  
XX diseases such as Alzheimer's disease, multiple sclerosis, schizophrenia,  
XX epilepsy and cancers. The present sequence is a protein encoded by one of  
XX the probes of the invention

RESULT 15

ABG52752

ID ABG52752 standard; peptide; 66 AA.

XX

AC ABG52752;

XX

DT 25-FEB-2003 (first entry)

XX

DE Human liver peptide, SEQ ID NO 31400.

XX

KW Human; liver; cirrhosis; hyperlipoproteinaemia; hyperlipidaemia;

KW hypercholesterolaemia; coronary heart disease.

XX

OS Homo sapiens.

XX

PN WO200157273-A2.

XX

PD 09-AUG-2001.

XX

PF 30-JAN-2001; 2001WO-US000664.

XX

PR 04-FEB-2000; 2000US-0180312P.

XX

PR 26-MAY-2000; 2000US-0207456P.

XX

PR 30-JUN-2000; 2000US-00608408.

XX

PR 03-AUG-2000; 2000US-00632366.

XX

PR 21-SEP-2000; 2000US-0234687P.

XX

PR 27-SEP-2000; 2000US-0236359P.

XX

PR 04-OCT-2000; 2000GB-00024263.

XX

PA (MOLE-) MOLECULAR DYNAMICS INC.

XX

PI Penn SG, Hanzel DK, Chen W, Rank DR;

XX

XX WPI; 2001-488898/53.

XX

XX Human genome-derived single exon nucleic acid probes useful for analyzing

XX gene expression in human adult liver.

XX

XX Claim 27; SEQ ID NO 31400; 658pp; English.

XX

XX The invention relates to a single exon nucleic acid probe (SENP) (I) for

XX measuring human gene expression in a sample derived from human adult

XX liver, comprising one of 13109 defined nucleotide sequences given in the

XX specification (or complements/ fragments). The probe hybridises at high

XX stringency to a nucleic acid molecule expressed in the human adult liver.

XX (I) may be used for predicting, measuring and displaying gene expression

XX in samples derived from human adult liver. The genes identified may be

XX involved in genetic liver diseases such as cirrhosis,

XX hyperlipoproteinaemia, hyperlipidaemia and hypercholesterolaemia which is

XX associated with coronary heart disease. ABG47348-ABG5930 represent human

XX liver single exon encoded peptides of the invention. Note: The sequence

XX information for this patent does not appear in the printed specification

XX but was obtained in electronic format directly from WIPO at

XX ftp.wipo.int/pub/published\_pct\_sequences

XX

XX Sequence 66 AA;

XX

Query Match 35.2%; Score 322; DB 4; Length 66;  
Best Local Similarity 100.0%; Pred. No. 2.2e-26;  
Matches 66; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 20 TTSQLPAFTTETIQEASEGPGADEVQVFAPANALPARSEAAAVQPVIGISQVRMNSKEK 79  
ID AAM58537 standard; protein; 66 AA.  
XX  
AC AAM58537;  
XX  
DT 05-NOV-2001 (first entry)  
XX  
DE Human brain expressed single exon probe encoded protein SEQ ID NO: 30642.  
XX  
KW Human; brain expressed exon; gene expression analysis; probe; microarray;  
KW Alzheimer's disease; multiple sclerosis; schizophrenia; epilepsy; cancer.  
XX  
OS Homo sapiens.  
XX  
PN WO200157275-A2.  
XX  
PD 09-AUG-2001.  
XX  
PF 30-JAN-2001; 2001WO-US000667.  
XX  
PR 04-FEB-2000; 2000US-0180312P.  
XX  
PR 26-MAY-2000; 2000US-0207456P.  
XX  
PR 30-JUN-2000; 2000US-00608408.  
XX  
PR 03-AUG-2000; 2000US-00632366.  
XX  
PR 21-SEP-2000; 2000US-0234687P.  
XX  
PR 27-SEP-2000; 2000US-0236359P.  
XX  
PR 04-OCT-2000; 2000GB-00024263.  
XX  
PA (MOLE-) MOLECULAR DYNAMICS INC.  
XX  
PI Penn SG, Hanzel DK, Chen W, Rank DR;  
XX  
XX WPI; 2001-488898/53.  
XX  
XX Human genome-derived single exon nucleic acid probes useful for analyzing

Search completed: March 17, 2004, 07:03:45  
Job time : 47.7739 secs

Search completed: March 17, 2004, 07:03:45  
Job time : 47.7739 secs

---

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 18, 2004, 13:12:33 ; Search time 32.5236 Seconds  
(without alignments)  
1425.213 Million cell updates/sec

Title: US-10-057-951-2\_COPY\_85\_263

Perfect score: 916

Sequence: 1 SGEAGVPEKRPCELDRCPET.....PVDPOEGSTPLMQAGTPGA 179

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1049977 seqs, 258955339 residues

Total number of hits satisfying chosen parameters: 1049977

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA:\*

- 1: /cgn2\_6/ptodata/2/pubpaa/US07\_PUBCOMB.pep.\*
- 2: /cgn2\_6/ptodata/2/pubpaa/PCT\_NEW\_PUB.pep.\*
- 3: /cgn2\_6/ptodata/2/pubpaa/US06\_NEW\_PUB.pep.\*
- 4: /cgn2\_6/ptodata/2/pubpaa/US06\_PUBCOMB.pep.\*
- 5: /cgn2\_6/ptodata/2/pubpaa/US07\_NEW\_PUB.pep.\*
- 6: /cgn2\_6/ptodata/2/pubpaa/PCTUS\_PUBCOMB.pep.\*
- 7: /cgn2\_6/ptodata/2/pubpaa/US08\_NEW\_PUB.pep.\*
- 8: /cgn2\_6/ptodata/2/pubpaa/US08\_PUBCOMB.pep.\*
- 9: /cgn2\_6/ptodata/2/pubpaa/US09\_PUBCOMB.pep.\*
- 10: /cgn2\_6/ptodata/2/pubpaa/US09\_PUBCOMB.pep.\*
- 11: /cgn2\_6/ptodata/2/pubpaa/US09C\_PUBCOMB.pep.\*
- 12: /cgn2\_6/ptodata/2/pubpaa/US09C\_PUBCOMB.pep.\*
- 13: /cgn2\_6/ptodata/2/pubpaa/US10\_PUBCOMB.pep.\*
- 14: /cgn2\_6/ptodata/2/pubpaa/US10\_PUBCOMB.pep.\*
- 15: /cgn2\_6/ptodata/2/pubpaa/US10C\_PUBCOMB.pep.\*
- 16: /cgn2\_6/ptodata/2/pubpaa/US10C\_PUBCOMB.pep.\*
- 17: /cgn2\_6/ptodata/2/pubpaa/US60\_NEW\_PUB.pep.\*
- 18: /cgn2\_6/ptodata/2/pubpaa/US60\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query	Score	Match	Length	ID	Description
1	916	100.0	263	9	US-09-084-491A-2	Sequence 2, Appli
2	916	100.0	263	13	US-10-102-704-2	Sequence 2, Appli
3	916	100.0	263	13	US-10-057-951-2	Sequence 2, Appli
4	913	99.7	263	14	US-10-210-951-44	Sequence 44, Appl
5	913	99.7	263	14	US-10-211-884-44	Sequence 44, Appl
6	322	35.2	66	9	US-09-864-761-38457	Sequence 38457, A
7	87.5	9.6	850	15	US-10-369-493-10111	Sequence 10111, A
8	86.5	9.4	507	15	US-10-289-762-174	Sequence 174, App
9	86.5	9.4	820	12	US-10-282-122A-54852	Sequence 54852, A
10	81	8.8	255	12	US-10-424-599-268014	Sequence 268014, A
11	80	8.7	1212	9	US-09-855-722-3	Sequence 3, Appli
12	80	8.7	1212	14	US-10-219-248-3	Sequence 3, Appli
13	80	8.7	1212	14	US-10-219-247-3	Sequence 3, Appli
14	80	8.7	1238	9	US-09-855-722-5	Sequence 5, Appli
15	80	8.7	1238	9	US-09-944-849-4	Sequence 4, Appli

16	80	8.7	1238	14	US-10-219-248-5	Sequence 5, Appli
17	80	8.7	1238	14	US-10-219-247-5	Sequence 5, Appli
18	79.5	8.6	769	15	US-10-320-800-10	Sequence 10, Appli
19	78.5	8.6	223	12	US-10-424-599-188306	Sequence 188306, Ap
20	78.5	8.6	2824	15	US-10-369-493-5166	Sequence 5166, Ap
21	78	8.5	152	12	US-10-424-599-210863	Sequence 210863, A
22	77.5	8.5	452	12	US-10-282-122A-70552	Sequence 70552, A
23	77.5	8.5	595	9	US-09-921-667-6	Sequence 6, Appli
24	77.5	8.5	1709	9	US-09-870-759-51	Sequence 51, Appli
25	77.5	8.5	1709	10	US-09-751-708A-51	Sequence 51, Appli
26	77.5	8.5	1709	10	US-09-863-776-58	Sequence 58, Appli
27	77.5	8.5	1709	10	US-09-863-776-60	Sequence 60, Appli
28	76.5	8.4	425	9	US-09-748-537-14	Sequence 14, Appli
29	76.5	8.4	425	9	US-09-821-831-2	Sequence 2, Appli
30	76.5	8.4	644	12	US-10-425-114-53819	Sequence 53819, A
31	76	8.3	83	15	US-10-346-000A-8	Sequence 8, Appli
32	76	8.3	490	15	US-10-369-493-21753	Sequence 8, Appli
33	76	8.3	755	14	US-10-156-761-11835	Sequence 21753, A
34	75.5	8.2	274	12	US-10-424-599-241849	Sequence 11835, A
35	75.5	8.2	276	9	US-09-864-761-46690	Sequence 241849, A
36	75	8.2	542	12	US-10-424-599-182444	Sequence 46690, A
37	75	8.2	595	9	US-09-826-312-9	Sequence 182444, A
38	75	8.2	595	9	US-09-935-727-11	Sequence 9, Appli
39	75	8.2	595	14	US-10-186-643-9	Sequence 11, Appli
40	75	8.2	595	14	US-10-207-655-154	Sequence 154, App
41	75	8.2	595	15	US-10-418-242-11	Sequence 11, Appli
42	75	8.2	788	14	US-10-156-761-14497	Sequence 14497, A
43	74.5	8.1	340	12	US-10-282-122A-48299	Sequence 48299, A
44	74.5	8.1	403	9	US-09-996-194-6	Sequence 6, Appli
45	74.5	8.1	412	14	US-10-161-572-63	Sequence 63, Appli

ALIGNMENTS

RESULT 1

US-09-084-491A-2  
; Sequence 2, Application US/09084491A  
; Patent No. US20020061576A1  
; GENERAL INFORMATION:  
; APPLICANT: MOORE, PAUL A.  
; APPLICANT: RUBEN, STEVEN M.  
; APPLICANT: EBER, REINHARD  
; TITLE OF INVENTION: TISSUE PLASMINOGEN ACTIVATOR-LIKE PROTEASE  
; NUMBER OF SEQUENCES: 16  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: HUMAN GENOME SCIENCES, INC.  
; STREET: 9410 KEY WEST AVENUE  
; CITY: ROCKVILLE  
; STATE: MD  
; COUNTRY: USA  
; ZIP: 20850  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent in Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/084,491A  
; FILING DATE: 27-MAY-1998  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: BROOKES, ANDERS A.  
; REGISTRATION NUMBER: 36,373  
; REFERENCE/DOCKET NUMBER: PF378  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (301) 309-8504  
; TELEFAX: (301) 309-8439  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 263 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear

```
; MOLECULE TYPE: protein
; US-09-084-491A-2

Query Match      100.0%; Score 916; DB 9; Length 263;
Best Local Similarity 100.0%; Pred. No. 9.8e-90;
Matches 179; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SGEAGVPEKPCEDLRCPETTSQALPAFTTEIQEASGPGADVQVFAPANALPARSEAA 60
DB 85 SGEAGVPEKPCEDLRCPETTSQALPAFTTEIQEASGPGADVQVFAPANALPARSEAA 144
QY 61 AVQPVIGISQVRVNSKEKDLGLTGLVGLGTTMMVILIIAGIILGYSYKRGKDLKEQH 120
DB 145 AVQPVIGISQVRVNSKEKDLGLTGLVGLGTTMMVILIIAGIILGYSYKRGKDLKEQH 204
QY 121 DQVCEREMQRIITPLSAFTNPTCEIVDEKTVVHTSQTVPDQEGSTPLMGAGTPGA 179
DB 205 DQVCEREMQRIITPLSAFTNPTCEIVDEKTVVHTSQTVPDQEGSTPLMGAGTPGA 263

RESULT 2
US-10-102-704-2
; Sequence 2, Application US/10102704
; Publication No. US20020164768A1
; GENERAL INFORMATION:
; APPLICANT: Moore et al.
; TITLE OF INVENTION: Tissue Plasminogen Activator-Like Protein
; FILE REFERENCE: PF378C1
; CURRENT APPLICATION NUMBER: US/10/102,704
; CURRENT FILING DATE: 2002-03-22
; PRIOR FILING DATE: 1998-05-27
; PRIOR APPLICATION NUMBER: 09/084,491
; PRIOR FILING DATE: 1998-05-27
; PRIOR APPLICATION NUMBER: 60/048,000
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: Patent in version 3.1
; SEQ ID NO 2
; LENGTH: 263
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-102-704-2

Query Match      100.0%; Score 916; DB 13; Length 263;
Best Local Similarity 100.0%; Pred. No. 9.8e-90;
Matches 179; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SGEAGVPEKPCEDLRCPETTSQALPAFTTEIQEASGPGADVQVFAPANALPARSEAA 60
DB 85 SGEAGVPEKPCEDLRCPETTSQALPAFTTEIQEASGPGADVQVFAPANALPARSEAA 144
QY 61 AVQPVIGISQVRVNSKEKDLGLTGLVGLGTTMMVILIIAGIILGYSYKRGKDLKEQH 120
DB 145 AVQPVIGISQVRVNSKEKDLGLTGLVGLGTTMMVILIIAGIILGYSYKRGKDLKEQH 204
QY 121 DQVCEREMQRIITPLSAFTNPTCEIVDEKTVVHTSQTVPDQEGSTPLMGAGTPGA 179
DB 205 DQVCEREMQRIITPLSAFTNPTCEIVDEKTVVHTSQTVPDQEGSTPLMGAGTPGA 263

RESULT 3
US-10-057-951-2
; Sequence 2, Application US/10057951
; Publication No. US20020177213A1
; GENERAL INFORMATION:
; APPLICANT: Moore et al.
; TITLE OF INVENTION: Tissue Plasminogen Activator-Like Protease
; FILE REFERENCE: PF378F1
; CURRENT APPLICATION NUMBER: US/10/057,951
; CURRENT FILING DATE: 2002-01-29
; PRIOR APPLICATION NUMBER: US 09/411,977
; PRIOR FILING DATE: 1999-10-04
; PRIOR APPLICATION NUMBER: US 09/084,491
; PRIOR FILING DATE: 1998-05-27
```

```
; PRIOR APPLICATION NUMBER: US 60/048,000
; PRIOR FILING DATE: 1997-05-28
; NUMBER OF SEQ ID NOS: 31
; SOFTWARE: Patent in Ver. 3.1
; SEQ ID NO 2
; LENGTH: 263
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-057-951-2

Query Match      100.0%; Score 916; DB 13; Length 263;
Best Local Similarity 100.0%; Pred. No. 9.8e-90;
Matches 179; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SGEAGVPEKPCEDLRCPETTSQALPAFTTEIQEASGPGADVQVFAPANALPARSEAA 60
DB 85 SGEAGVPEKPCEDLRCPETTSQALPAFTTEIQEASGPGADVQVFAPANALPARSEAA 144
QY 61 AVQPVIGISQVRVNSKEKDLGLTGLVGLGTTMMVILIIAGIILGYSYKRGKDLKEQH 120
DB 145 AVQPVIGISQVRVNSKEKDLGLTGLVGLGTTMMVILIIAGIILGYSYKRGKDLKEQH 204
QY 121 DQVCEREMQRIITPLSAFTNPTCEIVDEKTVVHTSQTVPDQEGSTPLMGAGTPGA 179
DB 205 DQVCEREMQRIITPLSAFTNPTCEIVDEKTVVHTSQTVPDQEGSTPLMGAGTPGA 263

RESULT 4
US-10-210-951-44
; Sequence 44, Application US/10210951
; Publication No. US20030170228A1
; GENERAL INFORMATION:
; APPLICANT: Ashkenazi, Avi J.
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Hillan, Kenneth J.
; APPLICANT: Marsters, Scot A.
; APPLICANT: Pan, James
; APPLICANT: Pitti, Robert M.
; APPLICANT: Roy, Margaret Ann
; APPLICANT: Smith, Victoria
; APPLICANT: Stone, Donna M.
; APPLICANT: Watanabe, Colin K.
; APPLICANT: Wood, William I.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE TREATMENT OF TUMOR
; FILE REFERENCE: P2931R1C1
; CURRENT APPLICATION NUMBER: US/10/210,951
; CURRENT FILING DATE: 2002-08-02
; PRIOR APPLICATION NUMBER: 60/014699
; PRIOR FILING DATE: 1996-04-01
; PRIOR APPLICATION NUMBER: 60/026943
; PRIOR FILING DATE: 1996-09-23
; PRIOR APPLICATION NUMBER: 60/059121
; PRIOR FILING DATE: 1997-07-17
; PRIOR APPLICATION NUMBER: 60/059352
; PRIOR FILING DATE: 1997-09-19
; PRIOR APPLICATION NUMBER: 60/062037
; PRIOR FILING DATE: 1997-10-10
; PRIOR APPLICATION NUMBER: 60/063755
; PRIOR FILING DATE: 1997-10-17
; PRIOR APPLICATION NUMBER: 60/063045
; PRIOR FILING DATE: 1997-10-24
; PRIOR APPLICATION NUMBER: 60/063046
; PRIOR FILING DATE: 1997-10-24
; PRIOR APPLICATION NUMBER: 60/066511
; PRIOR FILING DATE: 1997-11-24
; PRIOR APPLICATION NUMBER: 60/066772
; PRIOR FILING DATE: 1997-11-24
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 258
; SEQ ID NO 44
; LENGTH: 263
```



Matches 178; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 SGEAGVPEKPCEDLRCPETTSQALPAFTTEIQEASEGPGADEVQVFAPANALPARSEAA 60  
Db 85 SGEAGVPEKPCEDLRCPETTSQALPAFTTEIQEASEGPGADEVQVFAPANALPARSEAA 144

Qy 61 AVQPVIGISQVRVMSKSKDGLTGLVGLITMMVIIIAIGAGIILGYSYKRGKDLKEQH 120  
Db 145 AVQPVIGISQVRVMSKSKDGLTGLVGLITMMVIIIAIGAGIILGYSYKRGKDLKEQH 204

Qy 121 DQKVCEREMORIILPLSAFTNPCEIVDEKTVVHTSQTVPDQEGSTPLMQAGTPGA 179  
Db 205 DQKVCEREMORIILPLSAFTNPCEIVDEKTVVHTSQTVPDQEGSTPLMQAGTPGA 263

RESULT 6  
US-09-864-761-38457  
; Sequence 38457, Application US/09864761  
; Patent No. US20020048763A1  
; GENERAL INFORMATION:  
; APPLICANT: Penn, Sharon G.  
; APPLICANT: Rank, David R.  
; APPLICANT: Hanzel, David K.  
; APPLICANT: Chen, Wensheng  
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR  
; FILE REFERENCE: Aecmica-X-1  
; CURRENT APPLICATION NUMBER: US/09/864,761  
; CURRENT FILING DATE: 2001-05-23  
; PRIOR APPLICATION NUMBER: US 60/180,312  
; PRIOR FILING DATE: 2000-02-04  
; PRIOR APPLICATION NUMBER: US 60/207,456  
; PRIOR FILING DATE: 2000-05-26  
; PRIOR APPLICATION NUMBER: US 09/632,366  
; PRIOR FILING DATE: 2000-08-03  
; PRIOR APPLICATION NUMBER: GB 24263.6  
; PRIOR FILING DATE: 2000-10-04  
; PRIOR APPLICATION NUMBER: US 60/236,359  
; PRIOR FILING DATE: 2000-09-27  
; PRIOR APPLICATION NUMBER: PCT/US01/00666  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00667  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00664  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00669  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00665  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00668  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00663  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00662  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00661  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00670  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: US 60/234,687  
; PRIOR FILING DATE: 2000-09-21  
; PRIOR APPLICATION NUMBER: US 09/608,408  
; PRIOR FILING DATE: 2000-06-30  
; PRIOR APPLICATION NUMBER: US 09/774,203  
; PRIOR FILING DATE: 2001-01-29  
; NUMBER OF SEQ ID NOS: 49117  
; SOFTWARE: Annonmax Sequence Listing Engine vers. 1.1  
; SEQ ID NO 38457  
; LENGTH: 66  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; FEATURE:  
; OTHER INFORMATION: MAP TO AC002073.1

Query Match 99.7%; Score 913; DB 14; Length 263;  
Best Local Similarity 99.4%; Pred. No. 2.1e-89;  
Matches 178; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 SGEAGVPEKPCEDLRCPETTSQALPAFTTEIQEASEGPGADEVQVFAPANALPARSEAA 60  
Db 85 SGEAGVPEKPCEDLRCPETTSQALPAFTTEIQEASEGPGADEVQVFAPANALPARSEAA 144

Qy 61 AVQPVIGISQVRVMSKSKDGLTGLVGLITMMVIIIAIGAGIILGYSYKRGKDLKEQH 120  
Db 145 AVQPVIGISQVRVMSKSKDGLTGLVGLITMMVIIIAIGAGIILGYSYKRGKDLKEQH 204

Qy 121 DQKVCEREMORIILPLSAFTNPCEIVDEKTVVHTSQTVPDQEGSTPLMQAGTPGA 179  
Db 205 DQKVCEREMORIILPLSAFTNPCEIVDEKTVVHTSQTVPDQEGSTPLMQAGTPGA 263

RESULT 5  
US-10-211-884-44  
; Sequence 44, Application US/10211884  
; Publication No. US20030175900A1  
; GENERAL INFORMATION:  
; APPLICANT: Ashkenazi, Avi J.  
; APPLICANT: Goddard, Audrey  
; APPLICANT: Godowski, Paul J.  
; APPLICANT: Gurney, Austin L.  
; APPLICANT: Hillan, Kenneth J.  
; APPLICANT: Marsters, Scot A.  
; APPLICANT: Pan, James  
; APPLICANT: Pitti, Robert M.  
; APPLICANT: Roy, Margaret Ann  
; APPLICANT: Smith, Victoria  
; APPLICANT: Stone, Donna M.  
; APPLICANT: Watanabe, Colin K.  
; APPLICANT: Wood, William I.  
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE TREATMENT OF TUMOR  
; FILE REFERENCE: P2931R1C1  
; CURRENT APPLICATION NUMBER: US/10/211,884  
; CURRENT FILING DATE: 2002-08-02  
; PRIOR APPLICATION NUMBER: 60/014699  
; PRIOR FILING DATE: 1996-04-01  
; PRIOR APPLICATION NUMBER: 60/026943  
; PRIOR FILING DATE: 1996-09-23  
; PRIOR APPLICATION NUMBER: 60/059121  
; PRIOR FILING DATE: 1997-07-17  
; PRIOR APPLICATION NUMBER: 60/059352  
; PRIOR FILING DATE: 1997-09-19  
; PRIOR APPLICATION NUMBER: 60/062037  
; PRIOR FILING DATE: 1997-10-10  
; PRIOR APPLICATION NUMBER: 60/063755  
; PRIOR FILING DATE: 1997-10-17  
; PRIOR APPLICATION NUMBER: 60/063045  
; PRIOR FILING DATE: 1997-10-24  
; PRIOR APPLICATION NUMBER: 60/063046  
; PRIOR FILING DATE: 1997-10-24  
; PRIOR APPLICATION NUMBER: 60/066511  
; PRIOR FILING DATE: 1997-11-24  
; PRIOR APPLICATION NUMBER: 60/066772  
; PRIOR FILING DATE: 1997-11-24  
; Remaining Prior Application data removed - See File Wrapper or PALM.  
; NUMBER OF SEQ ID NOS: 258  
; SEQ ID NO 44  
; LENGTH: 263  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; US-10-211-884-44

Query Match 99.7%; Score 913; DB 14; Length 263;  
Best Local Similarity 99.4%; Pred. No. 2.1e-89;

```

; Publication No. US20040006218A1
; GENERAL INFORMATION:
; APPLICANT: Griffois, R.
; TITLE OF INVENTION: Chlamydia pneumoniae genomic sequence and polypeptides, fragment
; TITLE OF INVENTION: thereof and uses thereof, in particular for the diagnosis, prev
; TITLE OF INVENTION: and treatment of infection
; FILE REFERENCE: 9710-003-999
; CURRENT APPLICATION NUMBER: US/10/289,762
; CURRENT FILING DATE: 2003-03-27
; NUMBER OF SEQ ID NOS: 6849
; SEQ ID NO 174
; LENGTH: 507
; TYPE: PRT
; ORGANISM: Chlamydia pneumoniae
; FEATURE:
; NAME/KEY: SITE
; LOCATION: 1...507
; OTHER INFORMATION: Xaa-unknown or other
US-10-289-762-174

Query Match          9.4%; Score 86.5; DB 15; Length 507;
Best Local Similarity 24.1%; Pred. No. 2.4;
Matches 38; Conservative 28; Mismatches 49; Indels 43; Gaps 9;

QY 20 TTSQLAPAFTEIQE-----ASEGQADEVQVFAPANALPARSEAAA-VQPVIGIS 69
DB 235 TQSGSEAFETRLDTLLGVSLVFAPEHPDLSIV-----SPEQRDEVTAYYQESLRKS 288
QY 70 QRVRMS-REKKDLGTLY----VLGITMV-----IIIAIGAGIILGYSYKRGKDLKEQ 119
DB 289 ERDRISGVKTGTGFTGNVAKHITGNLLPWLISDYVVLGVGTGVVGV-----PA 339
QY 120 HDQKVCEREMQRTLPLSAFTNTCTEIVDEKTVVHTS 157
DB 340 HDER--DREFAEM-----FSLPIHEVIDDNGVCIIHN 369

RESULT 9
US-10-282-122A-54852
; Sequence 54852, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931

```

;; PRIOR FILING DATE: 2000-12-22  
;; PRIOR APPLICATION NUMBER: 60/267,636  
;; PRIOR FILING DATE: 2001-02-09  
;; PRIOR APPLICATION NUMBER: 60/269,308  
;; PRIOR FILING DATE: 2001-02-16  
;; Remaining Prior Application data removed - See File Wrapper or PALM.  
;; NUMBER OF SEQ ID NOS: 78614  
;; SOFTWARE: PatentIn version 3.1  
;; SEQ ID NO 54852  
;; LENGTH: 820  
;; TYPE: PRT  
;; ORGANISM: Chlamydia pneumoniae  
US-10-282-122A-54852  
  
Query Match 9.4%; Score 86.5; DB 12; Length 820;  
Best Local Similarity 24.1%; Pred. No. 4.8;  
Matches 38; Conservative 28; Mismatches 49; Indels 43; Gaps 9;  
  
QY 20 TTSQLPAFTTEIOE-----ASEGPGADEVQVPAPANALPARSEAAA-VQPVIGIS 69  
Db 235 TQEGSLAFTRLDTLLGVSLVIAPEHPDLDSIV-----SBEQDEVTAYVQESLRKS 288  
  
QY 70 QVRVMS-KEKDLGTLYG---VLGITMVV-----IIAIGAGILIGYSYKRGKDLKEQ 119  
Db 289 ERDRISVKTGTGFTGNAYAKHPITGNLLPWISDYVLGVTGVVGV-----PA 339  
  
QY 120 HDQKVCEREMQRIITLPSAFTNPTCEIVDEKTVVHTS 157  
Db 340 HDER--DREPAEM-----FSLPIHEVIDDNGVCIHNS 369  
  
RESULT 10  
US-10-424-599-268014  
; Sequence 268014, Application US/10424599  
; Publication No. US20040031072A1  
; GENERAL INFORMATION:  
; APPLICANT: La Rosa Thomas J  
; APPLICANT: Kovalic David K  
; APPLICANT: Zhou Yihua  
; APPLICANT: Cao Yongwei  
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With  
; FILE REFERENCE: 38-21(53223)B  
; CURRENT APPLICATION NUMBER: US/10/424,599  
; CURRENT FILING DATE: 2003-04-28  
; NUMBER OF SEQ ID NOS: 285684  
; SEQ ID NO 268014  
; LENGTH: 255  
; TYPE: PRT  
; ORGANISM: Glycine max  
; FEATURE:  
; OTHER INFORMATION: Clone ID: PAT\_MRT3847\_8403C.1.psp  
US-10-424-599-268014  
  
Query Match 8.8%; Score 81; DB 12; Length 255;  
Best Local Similarity 23.5%; Pred. No. 3.6;  
Matches 50; Conservative 27; Mismatches 66; Indels 70; Gaps 12;  
  
QY 7 PEKPCED-----LRCPETTS-----QALPAFTTEIOAS-----EGPGAD 42  
Db 31 PVRPLDSDVBYGCKTEGQHIIIVRYAECMGLFPRRIQSSRHOELGYKETQD-D 89  
  
QY 43 EYQVFAPANALPARSEAAVQPVIGIS-----QVRVMSK-EKDLGTLYG----- 87  
Db 90 EYEDL---CILLREVRQIPSVSAVSSGAIASDYQRLRVESVCSRLGLVSLVYLWKQDQ 145  
  
QY 88 -----VLGITMVIIAIGAGILG-----YSYKRGKDLKEQHDQKVCER- 127  
Db 146 SLLLOEMINGVAVTVKVAAMGLDPAXHLGKELAFNAYLHK-----LKEYINGVNCGE 201  
  
QY 128 -EMQRTITLPSAFTNPTCEIVDEKTVVHTSQT 159  
Db 202 GEVEITLDCPLFSNARI-VLDEYQVWVWSSDS 233

RESULT 11  
US-09-855-722-3  
; Sequence 3, Application US/09855722  
; Patent No. US20020049306A1  
; GENERAL INFORMATION:  
; APPLICANT: Sakano, Seiji  
; APPLICANT: Itoh, Akira  
; TITLE OF INVENTION: DIFFERENTIATION-SUPPRESSIVE POLYPEPTIDE  
; FILE REFERENCE: KP-8576  
; CURRENT APPLICATION NUMBER: US/09/855,722  
; CURRENT FILING DATE: 2001-05-16  
; PRIOR APPLICATION NUMBER: 09/214,278  
; PRIOR FILING DATE: 1999-01-26  
; NUMBER OF SEQ ID NOS: 32  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 3  
; LENGTH: 1212  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-855-722-3  
  
Query Match 8.7%; Score 80; DB 9; Length 1212;  
Best Local Similarity 22.4%; Pred. No. 42;  
Matches 55; Conservative 20; Mismatches 77; Indels 94; Gaps 11;  
  
QY 2 GEAGV--PEKRPC-----EDLRC-----PETTS-----QALPAFTTEIO 33  
Db 921 GEGGAEEPPSTPCLPERSGHLNDCARLTLFNRDHPQGTGVAICSGIRSLPATRAVAR 980  
  
QY 34 E-----ASEGPGADEVOV--FAPANALPARS--EAAAVQPVIGISQVRVMSKEXKD 81  
Db 981 DRLLVLLCDRASSGASAVEVAFSPARDLPDSSLIQGAHAIVAAITQR----- 1030  
  
QY 82 LGTLGVYLGITVMVIIAIGAGILG-----YSYKRGKDLKE 118  
Db 1031 -GNSSILLAVTEVKVETVVTGSGTGLLVPLVCGAFSVLACVVLVWVTKRRK--- 1085  
  
QY 119 QHDQKVCEREMQRIIT-----LPLSAFTNPTCEIVDEKTVVHTSQTVPDQEGSTP 169  
Db 1086 -----ERERSLPRESANNQWAFAPINPIERPQGHKDVLYCKNFPTPPRRADEA 1138  
  
QY 170 LMGQAG 175  
Db 1139 LQGPAG 1144  
  
RESULT 12  
US-10-219-248-3  
; Sequence 3, Application US/10219248  
; Publication No. US20030022368A1  
; GENERAL INFORMATION:  
; APPLICANT: Sakano, Seiji  
; APPLICANT: Itoh, Akira  
; TITLE OF INVENTION: DIFFERENTIATION-SUPPRESSIVE POLYPEPTIDE  
; FILE REFERENCE: KP-8576  
; CURRENT APPLICATION NUMBER: US/10/219,248  
; CURRENT FILING DATE: 2002-08-16  
; PRIOR APPLICATION NUMBER: US/09/214,278  
; PRIOR FILING DATE: 1999-01-26  
; NUMBER OF SEQ ID NOS: 32  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 3  
; LENGTH: 1212  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-219-248-3  
  
Query Match 8.7%; Score 80; DB 14; Length 1212;  
Best Local Similarity 22.4%; Pred. No. 42;  
Matches 55; Conservative 20; Mismatches 77; Indels 94; Gaps 11;

QY 2 GEAGV--PEKRPC-----EDLRC-----PETTS-----QALPAFTTEIQ 33  
Db 921 GEGCAEPPSTPCLPGRSHLDNNCARLTLHFNRDHVPQGTTVGAICSGIRSLPATRAVAR 980  
QY 34 E-----ASEGPGADEVQV-FAPANALPARS--EAAAVQPVIGISQVRMNSKEKD 81  
Db 981 DRLLVLLCDRASSGASAVEVAVSFSPARDLPDSSLIOGAHAIVAAITQR-----1030  
QY 82 LGTLGVLGITMVIILIAIGAGIILG-----YSYKRGKDLKE 118  
Db 1031 -GNSLLLAATVEVKVETVVTGSGSTGLLVPLVCGAFSVLWACVLCVWVTRKRRK---1085  
QY 119 QHDQKVCEREMORIT-----LPLSAFTNPTCEIVDEKTVVHTSQTPVDPQEGSTP 169  
Db 1086 -----ERERSRLPREESANNQWAPLPIRNPPIERPGGHKDVLYQCKNFPPPRRADEA 1138  
QY 170 LMGQAG 175  
Db 1139 LFGPAG 1144

RESULT 13  
US-10-219-247-3  
; Sequence 3, Application US/10219247  
; Publication No. US20030032781A1  
; GENERAL INFORMATION:  
; APPLICANT: Sakano, Seiji  
; TITLE OF INVENTION: DIFFERENTIATION-SUPPRESSIVE POLYPEPTIDE  
; FILE REFERENCE: KP-8576  
; CURRENT APPLICATION NUMBER: US/10/219,247  
; PRIOR FILING DATE: 2002-08-16  
; PRIOR FILING DATE: 2001-05-16  
; PRIOR FILING DATE: 2001-05-16  
; PRIOR FILING DATE: 1999-01-26  
; NUMBER OF SEQ ID NOS: 32  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 3  
; LENGTH: 1212  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-219-247-3

Query Match 8.7%; Score 80; DB 14; Length 1212;  
Best Local Similarity 22.4%; Pred. No. 42;  
Matches 55; Conservative 20; Mismatches 77; Indels 94; Gaps 11;  
QY 2 GEAGV--PEKRPC-----EDLRC-----PETTS-----QALPAFTTEIQ 33  
Db 921 GEGCAEPPSTPCLPGRSHLDNNCARLTLHFNRDHVPQGTTVGAICSGIRSLPATRAVAR 980  
QY 34 E-----ASEGPGADEVQV-FAPANALPARS--EAAAVQPVIGISQVRMNSKEKD 81  
Db 981 DRLLVLLCDRASSGASAVEVAVSFSPARDLPDSSLIOGAHAIVAAITQR-----1030  
QY 82 LGTLGVLGITMVIILIAIGAGIILG-----YSYKRGKDLKE 118  
Db 1031 -GNSLLLAATVEVKVETVVTGSGSTGLLVPLVCGAFSVLWACVLCVWVTRKRRK---1085  
QY 119 QHDQKVCEREMORIT-----LPLSAFTNPTCEIVDEKTVVHTSQTPVDPQEGSTP 169  
Db 1086 -----ERERSRLPREESANNQWAPLPIRNPPIERPGGHKDVLYQCKNFPPPRRADEA 1138  
QY 170 LMGQAG 175  
Db 1139 LFGPAG 1144

RESULT 14  
US-09-855-722-5  
; Sequence 5, Application US/09855722  
; Patent No. US20020049306A1

; GENERAL INFORMATION:  
; APPLICANT: Sakano, Seiji  
; APPLICANT: Itoh, Akira  
; TITLE OF INVENTION: DIFFERENTIATION-SUPPRESSIVE POLYPEPTIDE  
; FILE REFERENCE: KP-8576  
; CURRENT APPLICATION NUMBER: US/09/855,722  
; CURRENT FILING DATE: 2001-05-16  
; PRIOR FILING DATE: 2001-05-16  
; PRIOR FILING DATE: 1999-01-26  
; NUMBER OF SEQ ID NOS: 32  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 5  
; LENGTH: 1238  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-855-722-5

Query Match 8.7%; Score 80; DB 9; Length 1238;  
Best Local Similarity 22.4%; Pred. No. 43;  
Matches 55; Conservative 20; Mismatches 77; Indels 94; Gaps 11;  
QY 2 GEAGV--PEKRPC-----EDLRC-----PETTS-----QALPAFTTEIQ 33  
Db 947 GEGCAEPPSTPCLPGRSHLDNNCARLTLHFNRDHVPQGTTVGAICSGIRSLPATRAVAR 1006  
QY 34 E-----ASEGPGADEVQV-FAPANALPARS--EAAAVQPVIGISQVRMNSKEKD 81  
Db 1007 DRLLVLLCDRASSGASAVEVAVSFSPARDLPDSSLIOGAHAIVAAITQR-----1056  
QY 82 LGTLGVLGITMVIILIAIGAGIILG-----YSYKRGKDLKE 118  
Db 1057 -GNSLLLAATVEVKVETVVTGSGSTGLLVPLVCGAFSVLWACVLCVWVTRKRRK---1111  
QY 119 QHDQKVCEREMORIT-----LPLSAFTNPTCEIVDEKTVVHTSQTPVDPQEGSTP 169  
Db 1112 -----ERERSRLPREESANNQWAPLPIRNPPIERPGGHKDVLYQCKNFPPPRRADEA 1164  
QY 170 LMGQAG 175  
Db 1165 LFGPAG 1170

RESULT 15  
US-09-944-849-4  
; Sequence 4, Application US/09944849  
; Patent No. US20020151487A1  
; GENERAL INFORMATION:  
; APPLICANT: Nickoloff, Brian  
; APPLICANT: Miele, Lucio  
; TITLE OF INVENTION: METHOD AND REAGENTS FOR EPITHELIAL BARRIER FORMATION AND TREATME  
; TITLE OF INVENTION: MALIGNANT AND BENIGN SKIN DISORDERS BY MODULATING THE NOTCH PAI  
; FILE REFERENCE: 212583  
; CURRENT APPLICATION NUMBER: US/09/944,849  
; CURRENT FILING DATE: 2001-08-31  
; PRIOR FILING DATE: 2001-08-31  
; PRIOR FILING DATE: 2000-08-31  
; NUMBER OF SEQ ID NOS: 18  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 4  
; LENGTH: 1238  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-944-849-4

Query Match 8.7%; Score 80; DB 9; Length 1238;  
Best Local Similarity 22.4%; Pred. No. 43;  
Matches 55; Conservative 20; Mismatches 77; Indels 94; Gaps 11;  
QY 2 GEAGV--PEKRPC-----EDLRC-----PETTS-----QALPAFTTEIQ 33  
Db 947 GEGCAEPPSTPCLPGRSHLDNNCARLTLHFNRDHVPQGTTVGAICSGIRSLPATRAVAR 1006  
QY 34 E-----ASEGPGADEVQV-FAPANALPARS--EAAAVQPVIGISQVRMNSKEKD 81

Search completed: March 18, 2004, 13:24:45  
Job time : 32.5236 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 17, 2004, 07:00:39 ; Search time 13.2503 Seconds  
(without alignments)  
697.420 Million cell updates/sec

Title: US-10-057-951-2\_COPY\_85\_263

Perfect score: 916  
Sequence: 1 SGEAGVPEKPCEDLRCPET.....PVDQSGSTPLMGQAGTPGA 179

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents AA: \*  
1: /cgn2\_6/ptodata/2/iaa/5A COMB.pep.\*  
2: /cgn2\_6/ptodata/2/iaa/5B COMB.pep.\*  
3: /cgn2\_6/ptodata/2/iaa/6A COMB.pep.\*  
4: /cgn2\_6/ptodata/2/iaa/6B COMB.pep.\*  
5: /cgn2\_6/ptodata/2/iaa/PTCUT COMB.pep.\*  
6: /cgn2\_6/ptodata/2/iaa/backfiles1.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	916	100.0	263	US-09-411-977-2	Sequence 2, Appli
2	86.5	9.4	507	US-09-198-452A-174	Sequence 174, App
3	80	8.7	1212	US-09-214-278-3	Sequence 3, Appli
4	80	8.7	1212	US-09-855-722-3	Sequence 3, Appli
5	80	8.7	1238	US-09-214-278-5	Sequence 5, Appli
6	80	8.7	1238	US-09-855-722-5	Sequence 5, Appli
7	77.5	8.5	463	US-09-134-001C-3973	Sequence 3973, Ap
8	77.5	8.5	595	US-08-225-989-2	Sequence 2, Appli
9	77.5	8.5	595	US-08-570-923-2	Sequence 2, Appli
10	77.5	8.5	595	US-08-580-014-2	Sequence 2, Appli
11	77.5	8.5	595	US-09-079-785-2	Sequence 2, Appli
12	77.5	8.5	595	US-09-921-667-6	Sequence 6, Appli
13	77.5	8.5	595	US-09-628-126-2	Sequence 2, Appli
14	76.5	8.4	425	US-09-748-537-14	Sequence 14, Appli
15	75.5	8.2	219	US-09-134-001C-5651	Sequence 5651, Ap
16	75	8.2	595	US-08-232-087A-2	Sequence 2, Appli
17	75	8.2	595	US-09-006-353A-9	Sequence 9, Appli
18	75	8.2	595	US-09-573-986-9	Sequence 9, Appli
19	75	8.2	950	US-09-252-991A-29012	Sequence 29012, A
20	74.5	8.1	587	US-08-398-008A-2	Sequence 2, Appli
21	74.5	8.1	587	US-08-893-333-2	Sequence 2, Appli
22	74	8.1	1148	US-08-882-046-4	Sequence 4, Appli
23	73.5	8.0	1256	US-09-107-532A-4208	Sequence 4208, Ap
24	73.5	8.0	1023	US-09-762-724-14	Sequence 14, Appli
25	73	8.0	447	US-09-205-258-408	Sequence 408, App
26	73	8.0	1200	US-09-252-991A-31014	Sequence 31014, A
27	72.5	7.9	225	US-09-134-001C-3215	Sequence 3215, Ap

28	72.5	7.9	400	1	US-08-351-473B-5	Sequence 5, Appli
29	72.5	7.9	400	3	US-08-450-962-4	Sequence 4, Appli
30	72.5	7.9	400	3	US-08-450-962-6	Sequence 6, Appli
31	72.5	7.9	400	4	US-08-848-831-4	Sequence 4, Appli
32	72.5	7.9	400	4	US-08-848-831-6	Sequence 6, Appli
33	72	7.9	1780	1	US-08-769-309A-5	Sequence 5, Appli
34	72	7.9	1780	3	US-08-994-570-5	Sequence 5, Appli
35	71.5	7.8	707	4	US-09-228-986-80	Sequence 80, Appli
36	71	7.8	452	4	US-09-252-991A-31873	Sequence 31873, A
37	71	7.8	571	4	US-08-489-039A-14334	Sequence 14334, A
38	71	7.8	652	2	US-08-751-305-2	Sequence 2, Appli
39	70.5	7.7	1182	4	US-08-287-354-6	Sequence 6, Appli
40	70	7.6	249	4	US-09-252-991A-32233	Sequence 32233, A
41	70	7.6	377	4	US-09-489-039A-9429	Sequence 9429, Ap
42	69.5	7.6	332	4	US-09-489-039A-11558	Sequence 11558, A
43	69.5	7.6	1025	2	US-08-304-309-4	Sequence 4, Appli
44	69.5	7.6	1025	3	US-08-991-942-4	Sequence 4, Appli
45	69	7.5	239	4	US-09-489-039A-10522	Sequence 10522, A

ALIGNMENTS

RESULT 1

US-09-411-977-2

; Sequence 2, Application US/09411977

; Patent No. 6372473

; GENERAL INFORMATION:

; APPLICANT: Moore, Paul A.

; APPLICANT: Ruben, Steven M.

; APPLICANT: Ebner, Reinhard

; TITLE OF INVENTION: Tissue Plasminogen Activator-Like Protease

; FILE REFERENCE: PF3781

; CURRENT APPLICATION NUMBER: US/09/411,977

; EARLIER FILING DATE: 1999-10-04

; EARLIER APPLICATION NUMBER: 09/084,491

; EARLIER FILING DATE: 1998-05-27

; EARLIER APPLICATION NUMBER: 60/048,000

; EARLIER FILING DATE: 1997-05-28

; NUMBER OF SEQ ID NOS: 30

; SOFTWARE: Patent in Ver. 2.0

; SEQ ID NO 2

; TYPE: PRT

; ORGANISM: Homo sapiens

US-09-411-977-2

Query Match 100.0%; Score 916; DB 4; Length 263;

Best Local Similarity 100.0%; Pred. No. 1e-97;

Matches 179; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1	SGEAGVPEKPCEDLRCPETTSQALPAFTTEIQEASGPGADEVQVFAPANALPARGEAA	60
DB	85	SGEAGVPEKPCEDLRCPETTSQALPAFTTEIQEASGPGADEVQVFAPANALPARGEAA	144
QY	61	AVQPVIGISQVRNNSKEKDLGTLGTVLGTMMVIIAIGAILGYSYRGKDLKEQH	120
DB	145	AVQPVIGISQVRNNSKEKDLGTLGTVLGTMMVIIAIGAILGYSYRGKDLKEQH	204
QY	121	DQKVCEREMORITPLSAFTNPTCEIVDEKVVVHTSTPTVDPOEGSTPLMGQAGTPGA	179
DB	205	DQKVCEREMORITPLSAFTNPTCEIVDEKVVVHTSTPTVDPOEGSTPLMGQAGTPGA	263

RESULT 2

US-09-198-452A-174

; Sequence 174, Application US/09198452A

; Patent No. 6559294

; GENERAL INFORMATION:

; APPLICANT: Griffrath, R.

; TITLE OF INVENTION: Chlamydia pneumoniae genomic sequence and polypeptides, fragment

; TITLE OF INVENTION: thereof and uses thereof, in particular for the diagnosis, prev

; TITLE OF INVENTION: and treatment of infection

```
; FILE REFERENCE: 9710-003-999
; CURRENT APPLICATION NUMBER: US/09/198,452A
; CURRENT FILING DATE: 1998-11-24
; NUMBER OF SEQ ID NOS: 6849
; SEQ ID NO 174
; LENGTH: 507
; TYPE: PRT
; ORGANISM: Chlamydia pneumoniae
; FEATURE:
; NAME/KEY: SITE
; LOCATION: 1...507
; OTHER INFORMATION: Xaa-unknown or other
US-09-198-452A-174

Query Match
Best Local Similarity 9.4%; Score 86.5; DB 4; Length 507;
Matches 38; Conservative 28; Mismatches 49; Indels 43; Gaps 9;

QY 20 TTSQALPAFTTEIQE-----ASEGPGADEVQVPAPANALPARSERAAA-VQPVIGIS 69
DQ 235 TQGSLEAFTRDITLLGVSLVIAPEHPDLSIV-----SEEQDEVTAVQESLRKS 288
QY 70 QRVNMS-KEKDLGLGY-----VLGITMMV-----IIAIGAGIILGYSYKRGKDLKEQ 119
DQ 289 ERDRISSVKTKGFTGNVAKHEITGNLLPVWISDYVLGYGTGVVMGV-----PA 339
QY 120 HDQKVCEREMQRTLPLSAFTNPTCEIVDEKTVVHTS 157
DQ 340 HDER--DREPAEM-----FSLPIHEVIDDGVCIHSN 369

RESULT 3
US-09-214-278-3
; Sequence 3, Application US/09214278
; Patent No. 6291210
; GENERAL INFORMATION:
; APPLICANT: Sakano, Seiji
; APPLICANT: Itoh, Akira
; TITLE OF INVENTION: DIFFERENTIATION-SUPPRESSIVE POLYPEPTIDE
; FILE REFERENCE: KP-8576
; CURRENT APPLICATION NUMBER: US/09/214,278
; CURRENT FILING DATE: 1999-01-26
; NUMBER OF SEQ ID NOS: 32
; SOFTWARE: Patent in Ver. 2.1
; SEQ ID NO 3
; LENGTH: 1212
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-214-278-3

Query Match
Best Local Similarity 8.7%; Score 80; DB 3; Length 1212;
Matches 55; Conservative 20; Mismatches 77; Indels 94; Gaps 11;

QY 2 GEAGV--PEKRPC-----EDLRC-----PETTS-----QALPAFTTEIQ 33
DQ 921 GEGAAEPPPTCLPRSGHLDNNCARLTHFNDRHVPQGTIVCAICSGIRSLPATRAVAR 980
QY 34 E-----ASEGPGADEVQV-FAPANALPARS--EAAAVQPVIGISQVRVNSKEKD 81
DQ 981 DRLVLLCDRASSGASAVEVAVSFSPARDLPDSSLIQGAHAIVAIAITQR----- 1030
QY 82 LGTLGVVLGITMMVILIAIGAGIILG-----YSYKRGKDLKE 118
DQ 1031 -GNSLLLAATEVAVETVVTGGSGTGLLPVLCGASVLMACVLCVWVWTRKRRK-- 1085
QY 119 QHDQKVCEREMQRTIT-----LPLSAFTNPTCEIVDEKTVVHTSQTTPVDPQSGSTP 169
DQ 1086 -----ERERSRLPRESANNQWAPLPIRNPIERPGHKDVLQYCKNFTPPPRRADEA 1138
QY 170 LMGQAG 175
DQ 1139 LPGAAG 1144

US-09-214-278-3
```

```
Query Match
Best Local Similarity 8.7%; Score 80; DB 3; Length 1212;
Matches 55; Conservative 20; Mismatches 77; Indels 94; Gaps 11;

QY 2 GEAGV--PEKRPC-----EDLRC-----PETTS-----QALPAFTTEIQ 33
DQ 921 GEGAAEPPPTCLPRSGHLDNNCARLTHFNDRHVPQGTIVCAICSGIRSLPATRAVAR 980
QY 34 E-----ASEGPGADEVQV-FAPANALPARS--EAAAVQPVIGISQVRVNSKEKD 81
DQ 981 DRLVLLCDRASSGASAVEVAVSFSPARDLPDSSLIQGAHAIVAIAITQR----- 1030
QY 82 LGTLGVVLGITMMVILIAIGAGIILG-----YSYKRGKDLKE 118
DQ 1031 -GNSLLLAATEVAVETVVTGGSGTGLLPVLCGASVLMACVLCVWVWTRKRRK-- 1085
QY 119 QHDQKVCEREMQRTIT-----LPLSAFTNPTCEIVDEKTVVHTSQTTPVDPQSGSTP 169
DQ 1086 -----ERERSRLPRESANNQWAPLPIRNPIERPGHKDVLQYCKNFTPPPRRADEA 1138
QY 170 LMGQAG 175
DQ 1139 LPGAAG 1144

US-09-214-278-3
```

```
RESULT 4
US-09-855-722-3
; Sequence 3, Application US/09855722
; Patent No. 6638741
; GENERAL INFORMATION:
; APPLICANT: Sakano, Seiji
; APPLICANT: Itoh, Akira
; TITLE OF INVENTION: DIFFERENTIATION-SUPPRESSIVE POLYPEPTIDE
; FILE REFERENCE: KP-8576
; CURRENT APPLICATION NUMBER: US/09/855,722
; CURRENT FILING DATE: 2001-05-16
; PRIOR APPLICATION NUMBER: 09/214,278
; PRIOR FILING DATE: 1999-01-26
; NUMBER OF SEQ ID NOS: 32
; SOFTWARE: Patent in Ver. 2.1
; SEQ ID NO 3
; LENGTH: 1212
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-855-722-3

Query Match
Best Local Similarity 8.7%; Score 80; DB 4; Length 1212;
Matches 55; Conservative 20; Mismatches 77; Indels 94; Gaps 11;

QY 2 GEAGV--PEKRPC-----EDLRC-----PETTS-----QALPAFTTEIQ 33
DQ 921 GEGAAEPPPTCLPRSGHLDNNCARLTHFNDRHVPQGTIVCAICSGIRSLPATRAVAR 980
QY 34 E-----ASEGPGADEVQV-FAPANALPARS--EAAAVQPVIGISQVRVNSKEKD 81
DQ 981 DRLVLLCDRASSGASAVEVAVSFSPARDLPDSSLIQGAHAIVAIAITQR----- 1030
QY 82 LGTLGVVLGITMMVILIAIGAGIILG-----YSYKRGKDLKE 118
DQ 1031 -GNSLLLAATEVAVETVVTGGSGTGLLPVLCGASVLMACVLCVWVWTRKRRK-- 1085
QY 119 QHDQKVCEREMQRTIT-----LPLSAFTNPTCEIVDEKTVVHTSQTTPVDPQSGSTP 169
DQ 1086 -----ERERSRLPRESANNQWAPLPIRNPIERPGHKDVLQYCKNFTPPPRRADEA 1138
QY 170 LMGQAG 175
DQ 1139 LPGAAG 1144

US-09-214-278-5
```

```
Query Match
Best Local Similarity 8.7%; Score 80; DB 3; Length 1238;
Matches 55; Conservative 20; Mismatches 77; Indels 94; Gaps 11;

QY 2 GEAGV--PEKRPC-----EDLRC-----PETTS-----QALPAFTTEIQ 33
DQ 921 GEGAAEPPPTCLPRSGHLDNNCARLTHFNDRHVPQGTIVCAICSGIRSLPATRAVAR 980
QY 34 E-----ASEGPGADEVQV-FAPANALPARS--EAAAVQPVIGISQVRVNSKEKD 81
DQ 981 DRLVLLCDRASSGASAVEVAVSFSPARDLPDSSLIQGAHAIVAIAITQR----- 1030
QY 82 LGTLGVVLGITMMVILIAIGAGIILG-----YSYKRGKDLKE 118
DQ 1031 -GNSLLLAATEVAVETVVTGGSGTGLLPVLCGASVLMACVLCVWVWTRKRRK-- 1085
QY 119 QHDQKVCEREMQRTIT-----LPLSAFTNPTCEIVDEKTVVHTSQTTPVDPQSGSTP 169
DQ 1086 -----ERERSRLPRESANNQWAPLPIRNPIERPGHKDVLQYCKNFTPPPRRADEA 1138
QY 170 LMGQAG 175
DQ 1139 LPGAAG 1144

US-09-214-278-5
```

```
RESULT 5
US-09-214-278-5
; Sequence 5, Application US/09214278
; Patent No. 6291210
; GENERAL INFORMATION:
; APPLICANT: Sakano, Seiji
; APPLICANT: Itoh, Akira
; TITLE OF INVENTION: DIFFERENTIATION-SUPPRESSIVE POLYPEPTIDE
; FILE REFERENCE: KP-8576
; CURRENT APPLICATION NUMBER: US/09/214,278
; CURRENT FILING DATE: 1999-01-26
; NUMBER OF SEQ ID NOS: 32
; SOFTWARE: Patent in Ver. 2.1
; SEQ ID NO 5
; LENGTH: 1238
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-214-278-5

Query Match
Best Local Similarity 8.7%; Score 80; DB 3; Length 1238;
Matches 55; Conservative 20; Mismatches 77; Indels 94; Gaps 11;

QY 2 GEAGV--PEKRPC-----EDLRC-----PETTS-----QALPAFTTEIQ 33
DQ 921 GEGAAEPPPTCLPRSGHLDNNCARLTHFNDRHVPQGTIVCAICSGIRSLPATRAVAR 980
QY 34 E-----ASEGPGADEVQV-FAPANALPARS--EAAAVQPVIGISQVRVNSKEKD 81
DQ 981 DRLVLLCDRASSGASAVEVAVSFSPARDLPDSSLIQGAHAIVAIAITQR----- 1030
QY 82 LGTLGVVLGITMMVILIAIGAGIILG-----YSYKRGKDLKE 118
DQ 1031 -GNSLLLAATEVAVETVVTGGSGTGLLPVLCGASVLMACVLCVWVWTRKRRK-- 1085
QY 119 QHDQKVCEREMQRTIT-----LPLSAFTNPTCEIVDEKTVVHTSQTTPVDPQSGSTP 169
DQ 1086 -----ERERSRLPRESANNQWAPLPIRNPIERPGHKDVLQYCKNFTPPPRRADEA 1138
QY 170 LMGQAG 175
DQ 1139 LPGAAG 1144

US-09-214-278-5
```

```

Db 947 GCGAEEPPSTPCLPKSGHLNDCARLTLHFNRDHPVQGTGTAICSGIRSLPATRAVAR 1006
QY 34 E-----ASEGPGADEVOV--FAPANALPARS--EAAAVQPVIGISQVRMSKEKD 81
Db 1007 DRLVLLCDRASSGASAVEAVSFSPARDLPDSSLIQGAHAIVAAITQR----- 1056
QY 82 LGTLGYVLGTMVVIITAGAGIILG-----YSYKRGKDLKE 118
Db 1057 -GNSLLAVTEVKVETVTGSSSTGLLVPLVCGAFSVLWACVLCVWVTRKRRK---- 1111
QY 119 QHDQKVCEREMQRIIT-----LPLSAFTNPTCEIVDEKTVVHTSQTPVDPQSGSTP 169
Db 1112 -----ERERSRLPREESANNQWAPLPIRNPRIERPGRGHKDVLYQCKNFTPPRRADEA 1164
QY 170 LMGQAG 175
Db 1165 LPPGAG 1170

RESULT 6
US-09-855-722-5
; Sequence 5, Application US/09855722
; Patent No. 6638741
; GENERAL INFORMATION:
; APPLICANT: Sakano, Sei-ji
; APPLICANT: Itoh, Akira
; TITLE OF INVENTION: DIFFERENTIATION-SUPPRESSIVE POLYPEPTIDE
; FILE REFERENCE: KP-8576
; CURRENT APPLICATION NUMBER: US/09/855,722
; CURRENT FILING DATE: 2001-05-16
; PRIOR APPLICATION NUMBER: 09/214,278
; PRIOR FILING DATE: 1999-01-26
; NUMBER OF SEQ ID NOS: 32
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 5
; LENGTH: 1238
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-855-722-5

Query Match 8.7%; Score 80; DB 4; Length 1238;
Best Local Similarity 22.4%; Pred. No. 5.3;
Matches 55; Conservative 20; Mismatches 77; Indels 94; Gaps 11;

QY 2 GCGAV--PEKRPC-----EDLRC-----PETTS-----QALPAFTTEIQ 33
Db 947 GCGAEEPPSTPCLPKSGHLNDCARLTLHFNRDHPVQGTGTAICSGIRSLPATRAVAR 1006
QY 34 E-----ASEGPGADEVOV--FAPANALPARS--EAAAVQPVIGISQVRMSKEKD 81
Db 1007 DRLVLLCDRASSGASAVEAVSFSPARDLPDSSLIQGAHAIVAAITQR----- 1056
QY 82 LGTLGYVLGTMVVIITAGAGIILG-----YSYKRGKDLKE 118
Db 1057 -GNSLLAVTEVKVETVTGSSSTGLLVPLVCGAFSVLWACVLCVWVTRKRRK---- 1111
QY 119 QHDQKVCEREMQRIIT-----LPLSAFTNPTCEIVDEKTVVHTSQTPVDPQSGSTP 169
Db 1112 -----ERERSRLPREESANNQWAPLPIRNPRIERPGRGHKDVLYQCKNFTPPRRADEA 1164
QY 170 LMGQAG 175
Db 1165 LPPGAG 1170

RESULT 7
US-09-134-001C-3973
; Sequence 3973, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
```

```

; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 3973
; LENGTH: 463
; TYPE: PRT
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-3973

Query Match 8.5%; Score 77.5; DB 4; Length 463;
Best Local Similarity 31.1%; Pred. No. 2.4;
Matches 28; Conservative 14; Mismatches 37; Indels 11; Gaps 3;

QY 66 IGISQR--VRMSKEKDLGTGLYVLGTMVVIITAGAGIILGYSYKRGKDLKEQHDQ 122
Db 116 IGIRQLIMLDNRDNTSGTVKLTLEIVRTIFIEFTGALLAFYFYRDNPDLK----- 170
QY 123 KVCEREMQRIITLPLSAFTNPTCEIVDEKTV 152
Db 171 ---NALMQGIFVSVSATNNGLDITGESLV 197

RESULT 8
US-08-225-989-2
; Sequence 2, Application US/08225989
; Patent No. 5480981
; GENERAL INFORMATION:
; APPLICANT: Goodwin, Raymond G.
; APPLICANT: Smith, Craig A.
; APPLICANT: Armitage, Richard J.
; APPLICANT: Gruss, Hans-Jurgen
; TITLE OF INVENTION: No. 5480981el Cytokine That Binds CD30
; NUMBER OF SEQUENCES: 23
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Kathryn A. Seese, Immunex Corporation
; STREET: 51 University Street
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98101
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: Apple Macintosh
; OPERATING SYSTEM: Apple 7.1
; SOFTWARE: Microsoft Word, Version 5.1a
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/225,989
; FILING DATE: 12 APRIL 1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/966,775
; FILING DATE: 27-OCT-1992
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 907,224
; FILING DATE: 01-JUL-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 899,660
; FILING DATE: 15-JUN-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 892,459
; FILING DATE: 02-JUN-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 889,717
; FILING DATE: 26-MAY-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Seese, Kathryn A.
; REGISTRATION NUMBER: 32,172
```



```

; REFERENCE/DOCKET NUMBER: 2804-E
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206)587-0430
; TELEFAX: (206)233-0644
; TELEX: 756822
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 595 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-225-989-2

```

```

Query Match      8.58; Score 77.5; DB 1; Length 595;
Best Local Similarity 23.6%; Pred. No. 3.5; Indels 43; Gaps 11;
Matches 46; Conservative 32; Mismatches 32;

QY 4 AGVPEKPCEDLRCPETTSQALPAFT----TEIQEASEGPGADEVQVFAPANALPARSEA 59
Db 307 AGEVTKP-QDMAEKDITFEAPPLGTQDCNPTPENGAEPAFT-----SPTQSLLVDSQA 360

QY 60 AAVQPVIGISQVRNNSKEKDLGTLGVVLGITWVILIIAGAGIILGYSKRGKDLKEQ 119
Db 361 SKTLP-IPTAPVALSSTGKPVLDAGPVLFWVLVLVVVVGSAFL-----L 406

QY 120 HDQKVCEREMOR---ITLPLSAFTNPTCEIVDEKTVVHTSQT-----PVDPOEG 166
Db 407 CHRACRKRIRQKHLGCVFVQT-SQPKLELVDSRP---RRSTQLRSGASVTEPVAERG 462

QY 167 --STPLMQAGTPGA 179
Db 463 LMSQPLMETCHSVGA 477

```

```

RESULT 9
US-08-570-923-2
; Sequence 2, Application US/08570923
; Patent No. 5677430
; GENERAL INFORMATION:
; APPLICANT: Goodwin, Raymond G.
; APPLICANT: Smith, Craig A.
; APPLICANT: Armitage, Richard J.
; APPLICANT: Gruss, Hans-Jurgen
; TITLE OF INVENTION: No. 5677430el Cytokine That Binds CD30
; NUMBER OF SEQUENCES: 23
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Kathryn A. Seese, Immunex Corporation
; STREET: 51 University Street
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98101
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: Apple Macintosh
; OPERATING SYSTEM: Apple 7.1
; SOFTWARE: Microsoft Word, Version 5.1a
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/570,923
; FILING DATE: 12-DEC-1995
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/225,989
; FILING DATE: 12 APRIL 1994
; APPLICATION NUMBER: US 07/966,775
; FILING DATE: 27-OCT-1992
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 907,224
; FILING DATE: 01-JUL-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 899,660
; FILING DATE: 15-JUN-1992

```

```

; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 892,459
; FILING DATE: 02-JUN-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 889,717
; FILING DATE: 26-MAY-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Seese, Kathryn A.
; REGISTRATION NUMBER: 32,172
; REFERENCE/DOCKET NUMBER: 2804-E
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206)587-0430
; TELEFAX: (206)233-0644
; TELEX: 756822
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 595 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-570-923-2

```

```

Query Match      8.58; Score 77.5; DB 1; Length 595;
Best Local Similarity 23.6%; Pred. No. 3.5; Indels 43; Gaps 11;
Matches 46; Conservative 32; Mismatches 32;

QY 4 AGVPEKPCEDLRCPETTSQALPAFT----TEIQEASEGPGADEVQVFAPANALPARSEA 59
Db 307 AGEVTKP-QDMAEKDITFEAPPLGTQDCNPTPENGAEPAFT-----SPTQSLLVDSQA 360

QY 60 AAVQPVIGISQVRNNSKEKDLGTLGVVLGITWVILIIAGAGIILGYSKRGKDLKEQ 119
Db 361 SKTLP-IPTAPVALSSTGKPVLDAGPVLFWVLVLVVVVGSAFL-----L 406

QY 120 HDQKVCEREMOR---ITLPLSAFTNPTCEIVDEKTVVHTSQT-----PVDPOEG 166
Db 407 CHRACRKRIRQKHLGCVFVQT-SQPKLELVDSRP---RRSTQLRSGASVTEPVAERG 462

QY 167 --STPLMQAGTPGA 179
Db 463 LMSQPLMETCHSVGA 477

```

```

RESULT 10
US-08-580-014-2
; Sequence 2, Application US/08580014
; Patent No. 5753203
; GENERAL INFORMATION:
; APPLICANT: Goodwin, Raymond G.
; APPLICANT: Smith, Craig A.
; APPLICANT: Armitage, Richard J.
; APPLICANT: Gruss, Hans-Jurgen
; TITLE OF INVENTION: No. 5753203el Cytokine That Binds CD30
; NUMBER OF SEQUENCES: 23
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Kathryn A. Seese, Immunex Corporation
; STREET: 51 University Street
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98101
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: Apple Macintosh
; OPERATING SYSTEM: Apple 7.1
; SOFTWARE: Microsoft Word, Version 5.1a
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/580,014
; FILING DATE: 20-DEC-1995
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/225,989
; FILING DATE: 12 APRIL 1994

```

```
/ APPLICATION NUMBER: US 07/966,775
/ FILING DATE: 27-OCT-1992
/ CLASSIFICATION: 530
/ PRIOR APPLICATION DATA:
/ APPLICATION NUMBER: US 907,224
/ FILING DATE: 01-JUL-1992
/ PRIOR APPLICATION DATA:
/ APPLICATION NUMBER: US 899,660
/ FILING DATE: 15-JUN-1992
/ PRIOR APPLICATION DATA:
/ APPLICATION NUMBER: US 892,459
/ FILING DATE: 02-JUN-1992
/ PRIOR APPLICATION DATA:
/ APPLICATION NUMBER: US 889,717
/ FILING DATE: 26-MAY-1992
/ ATTORNEY/AGENT INFORMATION:
/ NAME: Seese, Kathryn A.
/ REGISTRATION NUMBER: 32,172
/ REFERENCE/DOCKET NUMBER: 2804-E
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: (206)587-0430
/ TELEFAX: (206)233-0644
/ TELEX: 756822
/ INFORMATION FOR SEQ ID NO: 2:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 595 amino acids
/ TYPE: amino acid
/ TOPOLOGY: linear
/ MOLECULE TYPE: protein
/ US-08-580-014-2

Query Match      8.5%; Score 77.5; DB 1; Length 595;
Best Local Similarity 23.6%; Pred. No. 3.5; Mismatches 32; Indels 43; Gaps 11;
Matches 46; Conservative 32;

QY 4 AGVPEKRCPELRCPETTSQALPAPT---TEIQASEGPGADEVQVFAPANALPARSEA 59
Db 307 AGETVTKP-QDMAEKDITFEAPPLGTQDCNPTPENGAPAST-----SPTQSLLVDSQA 360

QY 60 AAVQPVIGISQVRNMSKEKDLGTLGVVLGITWVILIAIGAGIILGYSKRGKDLKEQ 119
Db 361 SKTLP-ITSPAPVALSSTGKPVLDAGPVLFWILVVLVVVGSAFL-----L 406

QY 120 HDQKVCEREMQR---ITPLSAFTNPTCEIVDEKTVVHTSQT-----PVDPOEG 166
Db 407 CHRRACRKEIRQKHLCPVQT-SQPKLELVDSRP---RRSSTQLRSGASVTEPVAEBRG 462

QY 167 --STPLMQAGTPGA 179
Db 463 LMSQPLMETCHSVGA 477

RESULT 11
US-09-079-785-2
/ Sequence 2, Application US/09079785
/ Patent No. 6143869
/ GENERAL INFORMATION:
/ APPLICANT: Goodwin, Raymond G.
/ APPLICANT: Smith, Craig A.
/ APPLICANT: Armitage, Richard J.
/ APPLICANT: Gruss, Hans-Jurgen
/ TITLE OF INVENTION: No. 6143869el Cytokine That Binds CD30
/ NUMBER OF SEQUENCES: 23
/ CORRESPONDENCE ADDRESS:
/ ADDRESSES: Kathryn A. Seese, Immunex Corporation
/ STREET: 51 University Street
/ CITY: Seattle
/ STATE: Washington
/ COUNTRY: USA
/ ZIP: 98101
/ COMPUTER READABLE FORM:
/ MEDIUM TYPE: Floppy disk
/ COMPUTER: Apple Macintosh
```

```
/ OPERATING SYSTEM: Apple 7.1
/ SOFTWARE: Microsoft Word, Version 5.1a
/ CURRENT APPLICATION DATA:
/ APPLICATION NUMBER: US/09/079,785
/ FILING DATE:
/ CLASSIFICATION:
/ PRIOR APPLICATION DATA:
/ APPLICATION NUMBER: US 907,224
/ FILING DATE: 01-JUL-1992
/ PRIOR APPLICATION DATA:
/ APPLICATION NUMBER: US 899,660
/ FILING DATE: 15-JUN-1992
/ PRIOR APPLICATION DATA:
/ APPLICATION NUMBER: US 892,459
/ FILING DATE: 02-JUN-1992
/ PRIOR APPLICATION DATA:
/ APPLICATION NUMBER: US 889,717
/ FILING DATE: 26-MAY-1992
/ ATTORNEY/AGENT INFORMATION:
/ NAME: Seese, Kathryn A.
/ REGISTRATION NUMBER: 32,172
/ REFERENCE/DOCKET NUMBER: 2804-E
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: (206)587-0430
/ TELEFAX: (206)233-0644
/ TELEX: 756822
/ INFORMATION FOR SEQ ID NO: 2:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 595 amino acids
/ TYPE: amino acid
/ TOPOLOGY: linear
/ MOLECULE TYPE: protein
/ US-09-079-785-2

Query Match      8.5%; Score 77.5; DB 3; Length 595;
Best Local Similarity 23.6%; Pred. No. 3.5; Mismatches 32; Indels 43; Gaps 11;
Matches 46; Conservative 32;

QY 4 AGVPEKRCPELRCPETTSQALPAPT---TEIQASEGPGADEVQVFAPANALPARSEA 59
Db 307 AGETVTKP-QDMAEKDITFEAPPLGTQDCNPTPENGAPAST-----SPTQSLLVDSQA 360

QY 60 AAVQPVIGISQVRNMSKEKDLGTLGVVLGITWVILIAIGAGIILGYSKRGKDLKEQ 119
Db 361 SKTLP-ITSPAPVALSSTGKPVLDAGPVLFWILVVLVVVGSAFL-----L 406

QY 120 HDQKVCEREMQR---ITPLSAFTNPTCEIVDEKTVVHTSQT-----PVDPOEG 166
Db 407 CHRRACRKEIRQKHLCPVQT-SQPKLELVDSRP---RRSSTQLRSGASVTEPVAEBRG 462

QY 167 --STPLMQAGTPGA 179
Db 463 LMSQPLMETCHSVGA 477

RESULT 12
US-09-921-667-6
/ Sequence 6, Application US/09921667
/ Patent No. 6652854
/ GENERAL INFORMATION:
/ APPLICANT: Mohler, Kendall M.
/ APPLICANT: Barone, Dauphine S.
/ APPLICANT: Peschon, Jacques J.
/ APPLICANT: Kennedy, Mary K.
/ APPLICANT: Pluenneker, John D.
/ TITLE OF INVENTION: METHODS FOR TREATING AUTOIMMUNE AND CHRONIC INFLAMMATORY CONDITI
/ TITLE OF INVENTION: ANTAGONISTS OF CD30 OR CD30L
```

```
; FILE REFERENCE: 2959-A
; CURRENT APPLICATION NUMBER: US/09/921.667
; CURRENT FILING DATE: 2001-08-03
; PRIOR APPLICATION NUMBER: 60/224.079
; PRIOR FILING DATE: 2000-08-08
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: Patentin version 3.1
; SEQ ID NO 6
; LENGTH: 595
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-921-667-6

Query Match      8.5%; Score 77.5; DB 4; Length 595;
Best Local Similarity 23.6%; Pred. No. 3.5;
Matches 46; Conservative 32; Mismatches 74; Indels 43; Gaps 11;

QY 4 AGVPEKPCEDLRCPETTSQALPAFT----TEIQEASEGPGADEVQVFAPANALPARSEA 59
Db 307 AGEVTKP-QDMAEKDTTFEAPPLGTQDCNPTPENGEPAST-----SPTQSLLVDSQA 360
QY 60 AAVQPVIGISQVRVNSKEKDLGTLGVLGITMMVIIAIGAGIILGYSYKRGKDLKEQ 119
Db 361 SKTLF-IPTAPVALSSTGKPVLDG-GPVLFVILVILVVVVGSSAFL-----L 406
QY 120 HDQKVCEREMOR---ITLPLSAFTNPTCEIVDEKTVVHTSQT-----PVDPOEG 166
Db 407 CHRRACKRKIRQKHLCPVQVT-SQPKLEIVDSRP---RRSSTQLRSGASVTEPVAERG 462

RESULT 13
US-09-628-126-2
; Sequence 2, Application US/09628126
; Patent No. 6667039
; GENERAL INFORMATION:
; APPLICANT: Goodwin, Raymond G.
; APPLICANT: Smith, Craig A.
; APPLICANT: Armitage, Richard J.
; APPLICANT: Grues, Hans-Jurgen
; TITLE OF INVENTION: No. 6667039el Cytokine That Binds CD30
; NUMBER OF SEQUENCES: 23
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Kathryn A. Seese, Immunex Corporation
; STREET: 51 University Street
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98101
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: Apple 7.1
; SOFTWARE: Microsoft Word, Version 5.1a
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/628,126
; FILING DATE: 28-JULY-2000
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/570,923
; FILING DATE: 12-DEC-1995
; APPLICATION NUMBER: US/08/225,989
; FILING DATE: 12 APRIL 1994
; APPLICATION NUMBER: US 07/966,775
; FILING DATE: 27-OCT-1992
; APPLICATION NUMBER: US 907,224
; FILING DATE: 01-JUL-1992
; APPLICATION NUMBER: US 899,660
; FILING DATE: 15-JUN-1992
; APPLICATION NUMBER: US 892,459
```

```
; FILING DATE: 02-JUN-1992
; APPLICATION NUMBER: US 889,717
; FILING DATE: 26-MAY-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Seese, Kathryn A.
; REGISTRATION NUMBER: 32,172
; REFERENCE/DOCKET NUMBER: 2804-E
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206)587-0430
; TELEFAX: (206)233-0644
; TELEX: 756822
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 595 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-09-628-126-2

Query Match      8.5%; Score 77.5; DB 4; Length 595;
Best Local Similarity 23.6%; Pred. No. 3.5;
Matches 46; Conservative 32; Mismatches 74; Indels 43; Gaps 11;

QY 4 AGVPEKPCEDLRCPETTSQALPAFT----TEIQEASEGPGADEVQVFAPANALPARSEA 59
Db 307 AGEVTKP-QDMAEKDTTFEAPPLGTQDCNPTPENGEPAST-----SPTQSLLVDSQA 360
QY 60 AAVQPVIGISQVRVNSKEKDLGTLGVLGITMMVIIAIGAGIILGYSYKRGKDLKEQ 119
Db 361 SKTLF-IPTAPVALSSTGKPVLDG-GPVLFVILVILVVVVGSSAFL-----L 406
QY 120 HDQKVCEREMOR---ITLPLSAFTNPTCEIVDEKTVVHTSQT-----PVDPOEG 166
Db 407 CHRRACKRKIRQKHLCPVQVT-SQPKLEIVDSRP---RRSSTQLRSGASVTEPVAERG 462

RESULT 14
US-09-748-537-14
; Sequence 14, Application US/09748537
; Patent No. 6680167
; GENERAL INFORMATION:
; APPLICANT: Bertin, John
; APPLICANT: Chao, Moses V.
; TITLE OF INVENTION: NOVEL MOLECULES OF THE CARD-RELATED PROTEIN FAMILIE AND USES THEREOF
; FILE REFERENCE: 07334-316001
; CURRENT APPLICATION NUMBER: US/09/748,537
; CURRENT FILING DATE: 2000-12-26
; PRIOR APPLICATION NUMBER: US 09/099,041
; PRIOR FILING DATE: 1998-06-17
; PRIOR APPLICATION NUMBER: US 09/019,942
; PRIOR FILING DATE: 1998-02-06
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 14
; LENGTH: 425
; TYPE: PRT
; ORGANISM: Rattus rattus
US-09-748-537-14

Query Match      8.4%; Score 76.5; DB 4; Length 425;
Best Local Similarity 24.4%; Pred. No. 2.7;
Matches 42; Conservative 23; Mismatches 90; Indels 17; Gaps 5;

QY 3 EAGVPEKPCEDLRCPETTSQALPAFTTEIQEASEGPGADEVQVFAPANALPARSEAAAV 62
Db 175 ERQLRECTFWADAECEIIPGRWIPRSTPPEGSDDSTAPSTQEPFV-PPQDLVPSTVADMV 233
QY 63 QPVIGISQVRVNSKEKDLGTLGVLGITMMVIIAIGAGIILGYSYKRGKDLKEQHDQ 122
```

Db 234 TTVMGSSQPVVTR-----GTTDNLPV-YCSILAAVVVGLVAVIAFKWNSCK-QNKQ 284  
Qy 123 KVCEREMQRTITPLSAFTNPTCEIVDEKTVVHTSQTVPDPOEGSTPLMGQA 174  
Db 285 GANSR-----PVNQTPPEGEKLSHSDSGISVDSQSLHDQQTHTQTASGQA 329

RESULT 15  
US-09-134-001C-5651  
; Sequence 5651, Application US/09134001C  
; Patent No. 6380370  
; GENERAL INFORMATION:  
; APPLICANT: Lynn Doucette-Stamm et al  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS  
; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: GTC-007  
; CURRENT APPLICATION NUMBER: US/09/134,001C  
; CURRENT FILING DATE: 1998-08-13  
; PRIOR APPLICATION NUMBER: US 60/064,964  
; PRIOR FILING DATE: 1997-11-08  
; PRIOR APPLICATION NUMBER: US 60/055,779  
; PRIOR FILING DATE: 1997-08-14  
; NUMBER OF SEQ ID NOS: 5674  
; SEQ ID NO 5651  
; LENGTH: 219  
; TYPE: PRT  
; ORGANISM: Staphylococcus epidermidis  
US-09-134-001C-5651

Query Match 8.2% Score 75.5; DB 4; Length 219;  
Best Local Similarity 26.7%; Pred. No. 1.3;  
Matches 32; Conservative 22; Mismatches 33; Indels 33; Gaps 7;  
Qy 75 NSKEKKDL-GTLGYVLGITMMVIIAIGAGIILGYSY-----KRG-----113  
Db 7 NEREVNNMSGISKFISAIIVLLILLGLAFGI---YSFVDSKKGNERSLSDKTTQKEKKD 63  
Qy 114 -KDLKEQHDQKVCERE---MQRTIPLSAFTNPTCEIVDEKTVVHTSQTVP--DPOEG 166  
Db 64 DKDKKKKDKKSVBEKKNTCCTQQVPEQTQCTQCTV--QTPQRTTQTPVRQNPQTG 121

Search completed: March 17, 2004, 07:09:09  
Job time : 15.2503 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 17, 2004, 07:21:40 ; Search time 20 Seconds  
(without alignments)  
1163.917 Million cell updates/sec

Title: US-10-057-951-2\_COPY\_22\_263

Perfect score: 242

Sequence: 1 SGGCFWDNGHLYREDQTSFA.....FVDPQEGSTPLMGQATPGA 242

Scoring table:

Gapop 60.0 , Gapext 60.0

Searched: 283366 seqs, 96191526 residues

Word size : 0

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 100 summaries

Database :

PIR 78:.\*  
1: Pirl:.\*  
2: Pirl2:.\*  
3: Pirl3:.\*  
4: Pirl4:.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	9	3.7	331	2 G90037	hypothetical prote
2	8	3.3	429	2 S23258	adenylosuccinate s
3	8	3.3	602	2 G97293	ATP-dependent Zn p
4	8	3.3	716	1 A40332	macrophage-stimula
5	8	3.3	771	1 JC5061	probable DNA helic
6	8	3.3	771	2 C70716	hypothetical prote
7	8	3.3	3507	2 T34513	hypothetical prote
8	7	2.9	96	2 JC2561	chaperonin groESx
9	7	2.9	120	2 S69308	probable membrane
10	7	2.9	144	2 S74403	hypothetical prote
11	7	2.9	150	2 AC1060	conserved hypothet
12	7	2.9	154	2 F69520	conserved hypothet
13	7	2.9	164	2 A96594	hypothetical prote
14	7	2.9	222	2 C75342	conserved hypothet
15	7	2.9	226	2 D69780	hypothetical prote
16	7	2.9	238	2 S57610	granula associated
17	7	2.9	229	2 T40439	hypothetical prote
18	7	2.9	273	2 F67414	phage SfoI DNA pol
19	7	2.9	286	2 F89796	hypothetical prote
20	7	2.9	291	2 D90460	conserved hypothet
21	7	2.9	301	2 S57923	SEC14 protein - ye
22	7	2.9	303	2 T28999	hypothetical prote
23	7	2.9	314	2 T32247	hypothetical prote
24	7	2.9	326	2 AD2812	iron-sulfur cluste
25	7	2.9	326	2 B97394	hypothetical prote
26	7	2.9	332	2 AH1994	hypothetical prote
27	7	2.9	335	2 F64080	glycerol-3-phospha
28	7	2.9	335	2 D64397	hypothetical prote
29	7	2.9	346	2 AB2129	iron(III) dicitrat

30	7	2.9	356	2 A86590	hypothetical prote
31	7	2.9	356	2 D72033	hypothetical prote
32	7	2.9	355	2 F72033	hypothetical prote
33	7	2.9	365	2 B81505	hypothetical prote
34	7	2.9	365	2 C85590	hypothetical prote
35	7	2.9	388	2 S18560	xylyr protein - lac
36	7	2.9	398	2 B83252	probable amino aci
37	7	2.9	421	2 T43406	cullin-3 - fission
38	7	2.9	448	2 AB2740	biotin carboxylase
39	7	2.9	448	2 H97520	hypothetical prote
40	7	2.9	448	2 T06698	hypothetical prote
41	7	2.9	463	2 AD0799	probable membrane
42	7	2.9	483	2 A53918	chitinase (EC 3.2.
43	7	2.9	511	2 AB0397	multidrug resistan
44	7	2.9	512	2 G65048	multidrug resistan
45	7	2.9	512	2 D91072	hypothetical prote
46	7	2.9	512	2 G85916	multidrug resistan
47	7	2.9	512	2 AG0842	adenylosuccinate s
48	7	2.9	520	2 AI3295	tyrosine 3-monooxy
49	7	2.9	528	1 WHH074	urease (EC 3.5.1.5
50	7	2.9	567	1 D43719	hypothetical prote
51	7	2.9	585	2 S74477	phosphoribosylamin
52	7	2.9	591	2 S77707	phosphoribosylamin
53	7	2.9	592	2 S54489	phosphoribosylamin
54	7	2.9	629	2 C87048	probable ABC trans
55	7	2.9	631	2 T15370	hypothetical prote
56	7	2.9	701	2 C97310	ATP-dependent prot
57	7	2.9	710	1 I51283	hepatocyte growth
58	7	2.9	785	2 T38359	cullin 3 homolog -
59	7	2.9	788	2 S62405	hypothetical prote
60	7	2.9	937	2 A45082	neurotrophic recep
61	7	2.9	1101	2 G70951	probable ATP-depen
62	7	2.9	1157	1 S49247	parasporal crystal
63	7	2.9	1296	2 T16859	hypothetical prote
64	7	2.9	1538	2 H70846	hypothetical glyci
65	7	2.9	1638	2 T30313	chemotaxis protein
66	7	2.9	2472	2 E83594	still frameshift p
67	7	2.9	3433	1 GNVWVY	genome polypeptid
68	6	2.5	38	2 B86077	hypothetical prote
69	6	2.5	41	2 A42064	lactam utilization
70	6	2.5	49	2 B64323	hypothetical prote
71	6	2.5	69	2 H83236	cold acclimation p
72	6	2.5	71	2 B84284	hypothetical prote
73	6	2.5	73	2 AD1043	transcription regu
74	6	2.5	74	2 T14887	hypothetical prote
75	6	2.5	74	2 D82753	hypothetical prote
76	6	2.5	78	2 AG2814	30S ribosomal prot
77	6	2.5	81	2 E70774	probable atpE prot
78	6	2.5	81	2 T09579	h+transporting tw
79	6	2.5	82	2 A97820	hypothetical prote
80	6	2.5	83	1 C42645	ribosomal protein
81	6	2.5	83	2 F81664	ribosomal protein
82	6	2.5	85	2 G70661	hypothetical prote
83	6	2.5	86	2 H72054	ribosomal protein
84	6	2.5	86	2 C86570	S17 ribosomal prot
85	6	2.5	89	2 A60140	plasma (EC 3.4.21
86	6	2.5	93	2 AF1018	probable membrane
87	6	2.5	95	2 D64361	hypothetical prote
88	6	2.5	99	2 C70941	hypothetical prote
89	6	2.5	100	2 D87013	conserved hypothet
90	6	2.5	102	2 D90203	ATP synthase subun
91	6	2.5	105	2 D83243	hypothetical prote
92	6	2.5	106	1 R83243	acidic ribosomal p
93	6	2.5	106	2 T52147	ribosomal protein
94	6	2.5	106	2 F84797	hypothetical prote
95	6	2.5	107	2 E90976	hypothetical prote
96	6	2.5	107	2 C85823	hypothetical prote
97	6	2.5	108	2 H70128	hbv protein homol
98	6	2.5	108	2 T29957	hypothetical prote
99	6	2.5	109	1 T43825	ribosomal protein
100	6	2.5	109	1 R6BY11	acidic ribosomal p

Thu Mar 18 15:28:02 2004

ALIGNMENTS

RESULT 1

G90037  
hypothetical protein SA2162 [imported] - Staphylococcus aureus (strain N315)  
C:Species: Staphylococcus aureus  
C:Date: 10-May-2001 #sequence\_revision 10-May-2001 #text\_change 22-Oct-2001  
C:Accession: G90037  
R:Kuroda, M.; Ohta, T.; Uchiyama, I.; Baba, T.; Yuzawa, H.; Kobayashi, I.; Cui, L.; Oguc  
ma, A.; Mizutani-Ui, Y.; Kobayashi, N.; Sawano, T.; Inoue, R.; Kaito, C.; Sekimizu, K.;  
C.; Shiba, T.; Hattori, M.; Ogasawara, N.; Hayashi, H.; Hiramatsum, K.  
Lancet 357, 1225-1240, 2001  
A:Title: Whole genome sequencing of methicillin-resistant Staphylococcus aureus.  
A:Reference number: A89758; MUID:21311952; PMID:11418146  
A:Accession: G90037  
A>Status: Preliminary  
A:Molecule type: DNA  
A:Residues: 1-331 <KUR>  
A:Cross-references: GB:BA000018; PID:g13702323; PIDN:BA843464.1; GSPDB:GN00149  
A:Experimental source: strain N315  
C:Genetics:  
A:Gene: SA2162

Query Match 3.7%; Score 9; DB 2; Length 331;  
Best Local Similarity 100.0%; Pred.No.0.55; Mismatches 0; Indels 0; Gaps 0;  
Matches 9; Conservative 0

Qy 160 IIAIGAGII 168

Db 97 IIAIGAGII 105

RESULT 2

S23258  
adenylosuccinate synthase (EC 6.3.4.4) - Thiobacillus ferrooxidans  
N:Alternate names: IMP-aspartate ligase  
C:Species: Thiobacillus ferrooxidans  
C:Date: 22-Nov-1993 #sequence\_revision 10-Nov-1995 #text\_change 16-Jul-1999  
C:Accession: S23258  
R:Kusano, T.; Takeshima, T.; Sugawara, K.; Inoue, C.; Shiratori, T.; Yano, T.; Fukumori,  
J. Biol. Chem. 267, 11242-11247, 1992  
A:Title: Molecular cloning of the gene encoding Thiobacillus ferrooxidans Fe(II) oxidase  
A:Reference number: S23258; MUID:92283830; PMID:1317860  
A:Accession: S23258  
A>Status: Preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-429 <KUS>  
A:Cross-references: EMBL:X57324; NID:G48167; PIDN:CAA40593.1; PID:G48168  
A:Note: the nucleotide sequence was submitted to the EMBL Data Library, January 1991  
C:Genetics:  
A:Gene: purA  
C:Complex: homodimer

A:Function:  
A:Description: catalyzes the formation of AMP (with GDP and phosphate) from GTP, IMP, an  
A:Pathway: AMP biosynthesis; purine nucleotide biosynthesis (the first enzyme in the AMP  
C:Superfamily: adenylosuccinate synthase  
C:Keywords: AMP biosynthesis; GTP binding; homodimer; ligase; purine nucleotide biosynth  
F:141/Binding site: GTP (Lys) #status predicted

Query Match 3.3%; Score 8; DB 2; Length 429;  
Best Local Similarity 100.0%; Pred.No.7.2; Mismatches 0; Indels 0; Gaps 0;  
Matches 8; Conservative 0

Qy 148 LGYVLGIT 155

Db 260 LGYVLGIT 267

RESULT 3

G97293  
ATP-dependent Zn protease, FTSH [imported] - Clostridium acetobutylicum  
C:Species: Clostridium acetobutylicum

C:Date: 14-Sep-2001 #sequence\_revision 14-Sep-2001 #text\_change 30-Sep-2001  
C:Accession: G97293  
R:Nolling, J.; Bretton, G.; Omelchenko, M.V.; Markarova, K.S.; Zeng, Q.; Gibson, R.; Lee  
J.; Daly, M.J.; Bennett, G.N.; Koonin, E.V.; Smith, D.R.  
J. Bacteriol. 183, 4823-4838, 2001  
A:Title: Genome Sequence and Comparative Analysis of the Solvent-Producing Bacterium Cl  
A:Reference number: A96900; MUID:21359325; PMID:21359325  
A:Accession: G97293  
A>Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-602 <KUR>  
A:Cross-references: GB:AE001437; PIDN:AAK81138.1; PID:g15026273; GSPDB:GN00168  
A:Experimental source: Clostridium acetobutylicum ATCC824  
C:Genetics:  
A:Gene: CAC202  
C:Superfamily: cell division protein ftsH; FtsH/SEC18/CDC48-type ATP-binding domain hom

Query Match 3.3%; Score 8; DB 2; Length 602;  
Best Local Similarity 100.0%; Pred.No.9.6; Mismatches 0; Indels 0; Gaps 0;  
Matches 8; Conservative 0

Qy 63 VSGEAGVP 70

Db 213 VSGEAGVP 220

RESULT 4

A40332  
macrophage-stimulating protein 1 precursor - mouse  
N:Alternate names: hepatocyte growth factor-like protein  
C:Species: Mus musculus (house mouse)  
C:Date: 17-Jul-1992 #sequence\_revision 17-Jul-1992 #text\_change 18-Jun-1999  
C:Accession: A40332; B40332  
R:Degen, S.J.P.; Stuart, L.A.; Han, S.; Jamison, C.S.  
Biochemistry 30, 9781-9791, 1991  
A:Title: Characterization of the mouse cDNA and gene coding for a hepatocyte growth fac  
A:Reference number: A40332; MUID:92002017; PMID:1832957  
A:Accession: A40332  
A:Molecule type: DNA  
A:Residues: 1-716 <DSG>  
A:Cross-references: GB:M74180; NID:g193831; PIDN:AAA50166.1; PID:g193832  
A:Accession: B40332  
A:Molecule type: mRNA  
A:Residues: 1-18, P', 20-716 <DSG2>  
A:Cross-references: GB:M74181; NID:g193833; PIDN:AAA50167.1; PID:g193834

C:Genetics:  
A:Introns: 18/1; 67/2; 105/1; 143/2; 189/1; 229/2; 269/1; 334/2; 378/1; 412/2; 458/1; 4  
C:Complex: disulfide-bonded heterodimer of chains derived from the same precursor  
C:Superfamily: hepatocyte growth factor; kringle homology; trypsin homology  
C:Keywords: duplication; glycoprotein; growth factor; kringle  
F:1-31/Domain: signal sequence #status predicted <SIG>  
F:19-488, 489-716/Product: macrophage-stimulating protein 1 #status experimental <MAT>  
F:110-186/Domain: alpha chain #status experimental <ACH>  
F:191-268/Domain: kringle homology <KR1>  
F:292-370/Domain: kringle homology <KR2>  
F:378-457/Domain: kringle homology <KR3>  
F:484-711/Domain: beta chain #status experimental <BCH>  
F:489-709/Domain: trypsin homology <TR>  
F:72, 173, 305, 620/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 3.3%; Score 8; DB 1; Length 716;  
Best Local Similarity 100.0%; Pred.No.11; Mismatches 0; Indels 0; Gaps 0;  
Matches 8; Conservative 0

Qy 55 DPGPWCY 62

Db 163 DPGPWCY 170

RESULT 5

JCS061  
macrophage-stimulating protein 1 precursor - rat

C:Species: Rattus norvegicus (Norway rat)  
 C:Date: 31-Jan-1997 #sequence\_revision 31-Jan-1997 #text\_change 16-Jun-2000  
 C:Accession: JC5061  
 R;Onshiro, K.; Iwama, A.; Mateuno, K.; Ezaki, T.; Sakamoto, O.; Hameguchi, I.; Takasu, N. Biochem. Biophys. Res. Commun. 227, 273-280, 1996  
 A:Title: Molecular cloning of rat macrophage-stimulating protein and its involvement in A:Reference number: JC5061; MUID:97011126; PMID:8858136  
 A:Accession: JC5061  
 A:Molecule type: mRNA  
 A:Residues: 1-716 <OHS>  
 A:Cross-references: EMBL:X95096; NID:g1669718; PIDN:CA64473.1; PID:g1669719  
 C:Complex: disulfide-bonded heterodimer of chains derived from the same precursor  
 C:Superfamily: hepatocyte growth factor; kringle homology; trypsin homology  
 C:Keywords: duplication; glycoprotein; growth factor; kringle  
 F;1-31/Domain: signal sequence #status predicted <SIG>  
 F;32-488/Domain: macrophage-stimulating protein 1 #status predicted <MAT>  
 F;110-186/Domain: kringle homology <KRI1>  
 F;191-268/Domain: kringle homology <KRI2>  
 F;292-370/Domain: kringle homology <KRI3>  
 F;379-457/Domain: kringle homology <KRI4>  
 F;489-716/Domain: macrophage-stimulating protein 1 beta chain #status predicted <BCH>  
 F;489-709/Domain: trypsin homology <TRY>  
 F;72,305,620/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 3.3%; Score 8; DB 1; Length 716;  
 Best Local Similarity 100.0%; Pred. No. 11;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 55 DPGPWCY 62  
 |||||  
 Db 163 DPGPWCY 170

RESULT 6  
 C70716  
 probable DNA helicase - Mycobacterium tuberculosis (strain H37RV)  
 C:Species: Mycobacterium tuberculosis  
 C:Date: 17-Jul-1998 #sequence\_revision 17-Jul-1998 #text\_change 20-Jun-2000  
 C:Accession: C70716  
 R;Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S. Nature 393, 537-544, 1998  
 A:Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.  
 A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome A:Reference number: A70500; MUID:98295987; PMID:9634230  
 A:Accession: C70716  
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
 A:Molecule type: DNA  
 A:Residues: 1-771 <COL>  
 A:Cross-references: GB:279700; GB:AL123456; NID:g3261628; PIDN:CAB02001.1; PID:g1524213  
 A:Experimental source: strain H37RV  
 C:Genetics:  
 A:Gene: uvrd  
 C:Superfamily: helicase II

Query Match 3.3%; Score 8; DB 2; Length 771;  
 Best Local Similarity 100.0%; Pred. No. 12;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 37 SAPVSGAG 44  
 |||||  
 Db 692 SAPVSGAG 699

RESULT 7  
 T34513  
 hypothetical protein ZK783.1 - Caenorhabditis elegans  
 C:Species: Caenorhabditis elegans  
 C:Date: 29-Oct-1999 #sequence\_revision 29-Oct-1999 #text\_change 29-Oct-1999  
 C:Accession: T34513  
 R;Favella, A.; Vaudin, M.

submitted to the EMBL Data Library, August 1994  
 A:Description: The sequence of C. elegans cosmid ZK783.  
 A:Reference number: Z21536  
 A:Accession: T34513  
 A:Status: preliminary; translated from GB/EMBL/DDBJ  
 A:Molecule type: DNA  
 A:Residues: 1-3507 <PAV>  
 A:Cross-references: EMBL:U13646; PIDN:AAC24418.1; GSPDB:GN00021; CESP:ZK783.1  
 A:Experimental source: strain Bristol N2; clone ZK783  
 C:Genetics:  
 A:Gene: CESP:ZK783.1  
 A:Map position: 3  
 A:Introns: 14/1; 48/2; 84/1; 196/3; 303/1; 381/1; 586/1; 605/1; 1175/3; 1207/1; 1405/2; 3504/1

Query Match 3.3%; Score 8; DB 2; Length 3507;  
 Best Local Similarity 100.0%; Pred. No. 45;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 64 SGEAGVPE 71  
 |||||  
 Db 1269 SGEAGVPE 1276

RESULT 8  
 JC2561  
 chaperonin groES protein - Amoeba proteus  
 C:Species: Amoeba proteus  
 C:Date: 13-Jun-1995 #sequence\_revision 14-Jul-1995 #text\_change 26-Aug-1999  
 C:Accession: JC2561  
 R;Ahn, T.I.; Lim, S.T.; Leeu, H.K.; Lee, J.E.; Jeon, K.W.  
 Gene 148[128], 43-49, 1994  
 A:Title: A novel strong promoter of the groEx operon of symbiotic bacteria in Amoeba pr A:Reference number: JC2561  
 A:Note: due to a typographical error the volume number 148 appears as 128  
 A:Accession: JC2561  
 A:Molecule type: DNA  
 A:Residues: 1-96 <AHN>  
 A:Cross-references: GB:M86549; NID:g155400; PIDN:AAC09380.1; PID:g155401  
 C:Comment: This protein is involved in the assembly of oligomeric protein complexes, an C:Genetics:  
 A:Gene: groESX  
 C:Superfamily: chaperonin groES  
 C:Keywords: molecular chaperone

Query Match 2.9%; Score 7; DB 2; Length 96;  
 Best Local Similarity 100.0%; Pred. No. 20;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 160 IIAIGAG 166  
 |||||  
 Db 40 IIAIGAG 46

RESULT 9  
 S69308  
 probable membrane protein YLR302c - yeast (Saccharomyces cerevisiae)  
 N:Alternate names: hypothetical protein L8003.2-B  
 C:Species: Saccharomyces cerevisiae  
 C:Date: 20-Jul-1996 #sequence\_revision 23-Aug-1996 #text\_change 19-Apr-2002  
 C:Accession: S69308  
 R;Pauley, A.  
 submitted to the EMBL Data Library, November 1994  
 A:Description: The sequence of S. cerevisiae cosmid 8003.  
 A:Reference number: S50366  
 A:Accession: S69308  
 A:Molecule type: DNA  
 A:Residues: 1-120 <PAU>  
 A:Cross-references: EMBL:U17243; NID:g596030; PID:g2340968; GSPDB:GN00012; MIPS:YLR302c  
 C:Genetics:  
 A:Gene: MIPS:YLR302c  
 A:Cross-references: SGD:S0004293  
 A:Map position: 12R

C:Superfamily: Saccharomyces cerevisiae probable membrane protein YLR302c  
C:Keywords: transmembrane protein  
F:44-60/Domain: transmembrane #status predicted <TM>

Query Match 2.9%; Score 7; DB 2; Length 120;  
Best Local Similarity 100.0%; Pred. No. 25;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 208 EIVDEKT 214  
Db 25 EIVDEKT 31  
|||||

RESULT 10  
S74403  
hypothetical protein slr0491 - *Synechocystis* sp. (strain PCC 6803)  
C:Species: *Synechocystis* sp.  
A:Variety: PCC 6803  
C:Date: 25-Apr-1997 #sequence\_revision 25-Apr-1997 #text\_change 08-Oct-1999  
R:Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.; Miyajima, N.; O. K.; Okumura, S.; Shimpo, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yasuda  
DNA Res. 3, 109-136, 1996  
A:Title: Sequence analysis of the genome of the unicellular cyanobacterium *Synechocystis*  
S.  
A:Reference number: S74322; MUID:97061201; PMID:8905231  
A:Accession: S74403  
A:Status: nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-144 <KAN>  
A:Cross-references: EMBL:D64001; GB:AB001339; NID:G1001102; PIDN:BAAL0321.1; PID:G101097  
A:Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1996

Query Match 2.9%; Score 7; DB 2; Length 144;  
Best Local Similarity 100.0%; Pred. No. 29;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 147 TLGVVLG 153  
Db 40 TLGVVLG 46  
|||||

RESULT 11  
AC1060  
conserved hypothetical protein STY4806 [imported] - *Salmonella enterica* subsp. *enterica*  
C:Species: *Salmonella enterica* subsp. *enterica* serovar Typhi  
A:Note: this species has also been called *Salmonella typhi*  
C:Date: 09-Nov-2001 #sequence\_revision 09-Nov-2001 #text\_change 18-Nov-2002  
C:Accession: AC1060  
R:Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Churcher,  
th, T.; Cronin, A.; Davis, P.; Davies, R.M.; Dowd, L.; White, N.; Farrar,  
S.; Moule, S.; O'Gaora, P.  
Nature 413, 848-852, 2001  
A:Authors: Parry, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.;  
A:Title: Complete genome sequence of a multiple drug resistant *Salmonella enterica* serov  
A:Reference number: AB0502; MUID:21534947; PMID:11677608  
A:Accession: AC1060  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-150 <PAR>  
A:Cross-references: GB:AL513382; PIDN:CAD06928.1; PID:G16505576; GSFPDB:GN00176  
C:Genetics:  
A:Gene: STY4806  
C:Superfamily: hypothetical protein HI0227

Query Match 2.9%; Score 7; DB 2; Length 150;  
Best Local Similarity 100.0%; Pred. No. 30;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 210 VDEKTVV 216  
Db 107 VDEKTVV 113  
|||||

RESULT 12  
F69520  
conserved hypothetical protein AF2166 - *Archaeoglobus fulgidus*  
C:Species: *Archaeoglobus fulgidus*  
C:Date: 05-Dec-1997 #sequence\_revision 05-Dec-1997 #text\_change 22-Oct-1999  
C:Accession: F69520  
R:Klenk, H.P.; Clayton, R.A.; Tomb, J.F.; White, O.; Nelson, K.E.; Ketchum, K.A.; Dodso  
.; Fleischmann, R.D.; Quackenbush, J.; Lee, N.H.; Sutton, G.G.; Gill, S.; Kirkness, E.F  
Glodek, A.; Zhou, L.; Overbeek, R.; Gocayne, J.D.; Weidman, J.F.; McDonald, L.  
Nature 390, 364-370, 1997  
A:Authors: Uterback, T.; Cotton, M.D.; Spriggs, T.; Artiach, P.; Kaine, B.P.; Sykes, S  
Smith, H.O.; Woese, C.R.; Venter, J.C.  
A:Title: The complete genome sequence of the hyperthermophilic, sulfate-reducing archae  
A:Reference number: A69250; MUID:98049343; PMID:9389475  
A:Accession: F69520  
A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-154 <KLE>  
A:Cross-references: GB:AE000955; GB:AE000782; NID:G2689278; PIDN:AAE89089.1; PID:G26483

Query Match 2.9%; Score 7; DB 2; Length 154;  
Best Local Similarity 100.0%; Pred. No. 31;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 166 GIILGYS 172  
Db 104 GIILGYS 110  
|||||

RESULT 13  
A96594  
hypothetical protein F7A10.6 [imported] - *Arabidopsis thaliana*  
C:Species: *Arabidopsis thaliana* (mouse-ear cress)  
C:Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 31-Mar-2001  
C:Accession: A96594  
R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso  
Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.  
ansen, N.F.; Hughes, B.; Huizart, L.  
Nature 408, 816-820, 2000  
A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C  
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marziani  
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.  
A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,  
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.  
A:Title: Sequence and analysis of chromosome 1 of the plant *Arabidopsis*.  
A:Reference number: A86141; MUID:21016719; PMID:111130712  
A:Accession: A96594  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-164 <STO>  
A:Cross-references: GB:AE005173; NID:G10645434; PIDN:AAG21551.1; GSFPDB:GN00141  
C:Genetics:  
A:Gene: F7A10.6  
A:Map position: 1

Query Match 2.9%; Score 7; DB 2; Length 164;  
Best Local Similarity 100.0%; Pred. No. 32;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 142 KKDGLTL 148  
Db 145 KKDGLTL 151  
|||||

RESULT 14  
C75342  
conserved hypothetical protein - *Deinococcus radiodurans* (strain R1)  
C:Species: *Deinococcus radiodurans*  
C:Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 17-Mar-2000  
C:Accession: C75342  
R:White, O.; Eisen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J.;  
M.; Shen, M.; Vamathevan, J.J.; Lam, P.; McDonald, L.; Uterback, T.; Zalewski, C.; M

Query Match 2.9%; Score 7; DB 2; Length 164;  
Best Local Similarity 100.0%; Pred. No. 32;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 142 KKDGLTL 148  
Db 145 KKDGLTL 151  
|||||

RESULT 14  
C75342  
conserved hypothetical protein - *Deinococcus radiodurans* (strain R1)  
C:Species: *Deinococcus radiodurans*  
C:Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 17-Mar-2000  
C:Accession: C75342  
R:White, O.; Eisen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J.;  
M.; Shen, M.; Vamathevan, J.J.; Lam, P.; McDonald, L.; Uterback, T.; Zalewski, C.; M

Query Match 2.9%; Score 7; DB 2; Length 164;  
Best Local Similarity 100.0%; Pred. No. 32;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 142 KKDGLTL 148  
Db 145 KKDGLTL 151  
|||||

RESULT 14  
C75342  
conserved hypothetical protein - *Deinococcus radiodurans* (strain R1)  
C:Species: *Deinococcus radiodurans*  
C:Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 17-Mar-2000  
C:Accession: C75342  
R:White, O.; Eisen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J.;  
M.; Shen, M.; Vamathevan, J.J.; Lam, P.; McDonald, L.; Uterback, T.; Zalewski, C.; M

Query Match 2.9%; Score 7; DB 2; Length 164;  
Best Local Similarity 100.0%; Pred. No. 32;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 142 KKDGLTL 148  
Db 145 KKDGLTL 151  
|||||

RESULT 14  
C75342  
conserved hypothetical protein - *Deinococcus radiodurans* (strain R1)  
C:Species: *Deinococcus radiodurans*  
C:Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 17-Mar-2000  
C:Accession: C75342  
R:White, O.; Eisen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J.;  
M.; Shen, M.; Vamathevan, J.J.; Lam, P.; McDonald, L.; Uterback, T.; Zalewski, C.; M



S.; Smith, H.O.; Venter, J.C.; Fraser, C.M.  
Science 286, 1571-1577, 1999  
A:Title: Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1.  
A:Reference number: A75250; MUID:20036896; PMID:10567266  
A:Accession: C75342  
A>Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-222 <WHI>  
A:Cross-references: GB:AE002027; GB:AE00513; NID:g6459655; PIDN:AAF11429.1; PID:g645965  
A:Experimental source: strain R1  
C:Genetics:  
A:Gene: DR1875  
A:Map position: 1

Query Match 2.9%; Score 7; DB 2; Length 222;  
Best Local Similarity 100.0%; Pred. No. 42;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 35 LASAPVS 41  
|||||  
Db 2 LASAPVS 8

RESULT 15  
D69780  
hypothetical protein ydfP - Bacillus subtilis  
C:Species: Bacillus subtilis  
C:Date: 05-Dec-1997 #sequence\_revision 05-Dec-1997 #text\_change 15-Oct-1999  
C:Accession: D69780  
R:Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Berber  
C.; Bron, S.; Brouillet, S.; Brusch, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.; Ch  
A.; Ehrlich, S.D.; Emerson, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrari, E.  
Nature 390, 249-256, 1997  
A:Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Gallizzi, A.; Gall  
lech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hulio, M.F.  
Koetter, P.; Koningsstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardinois,  
A:Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Mausel  
Y, M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portetelle  
Rieger, M.; Rivolta, C.; Rocha, E.; Roche, M.; Rose, M.; Sadaie, Y.; Sato, T.; Scanlon  
A:Authors: Schleich, S.; Schroeter, R.; Scoffone, P.; Sekiguchi, J.; Sekowska, A.; Seron  
akeuchi, M.; Tamakoshi, A.; Tanaka, T.; Terpstra, P.; Tognoni, A.; Tosato, V.; Uchiyama,  
T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida, K  
A:Authors: Yoshikawa, H.F.; Zumbstein, E.; Yoshikawa, H.; Danchin, A.  
A:Title: The complete genome sequence of the Gram-positive bacterium Bacillus subtilis.  
A:Reference number: A69580; MUID:98044033; PMID:9384377  
A:Accession: D69780  
A>Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-226 <KUN>  
A:Cross-references: GB:Z99106; GB:AL009126; NID:g2632653; PIDN:CAB12346.1; PID:el182505;  
A:Experimental source: strain 168  
C:Genetics:  
A:Gene: ydfP

Query Match 2.9%; Score 7; DB 2; Length 226;  
Best Local Similarity 100.0%; Pred. No. 43;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 139 SXKKDL 145  
|||||  
Db 99 SXKKDL 105

RESULT 16  
S57610  
granula associated protein 24 - Alcaligenes eutrophus  
C:Species: Alcaligenes eutrophus  
C:Date: 19-Oct-1995 #sequence\_revision 03-Nov-1995 #text\_change 09-Sep-1997  
C:Accession: S57610  
R:Wieczorek, R.; Fries, A.; Steinbuechel, A.; Mayer, F.  
submitted to the EMBL Data Library, March 1995  
A:Description: Analysis of a 24 kDa protein associated with the polynhydroxyalkanoic acid  
A:Reference number: S57610

A:Accession: S57610  
A>Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-228 <WIB>  
A:Cross-references: EMBL:X85729; NID:g886423; PID:g886424

Query Match 2.9%; Score 7; DB 2; Length 228;  
Best Local Similarity 100.0%; Pred. No. 43;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 122 AAAYQPV 128  
|||||  
Db 69 AAAYQPV 75

RESULT 17  
T40439  
hypothetical protein SPBC409.12c - fission yeast (Schizosaccharomyces pombe)  
C:Species: Schizosaccharomyces pombe  
C:Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 03-Dec-1999  
C:Accession: T40439  
R:Lyne, M.H.; Rajandream, M.A.; Barrell, B.G.; Chillingworth, T.; Churcher, C.M.  
submitted to the EMBL Data Library, August 1999  
A:Reference number: 221929  
A:Accession: T40439  
A>Status: preliminary; translated from GB/EMBL/DDBJ  
A:Molecule type: DNA  
A:Residues: 1-229 <LYN>  
A:Cross-references: EMBL:AL109822; PIDN:CAB52614.1; GSPDB:GN00067; SPDB:SPBC409.12c  
A:Experimental source: strain 972h-; cosmid c409  
C:Genetics:  
A:Gene: SPDB:SPBC409.12c  
A:Map position: 2

Query Match 2.9%; Score 7; DB 2; Length 229;  
Best Local Similarity 100.0%; Pred. No. 43;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 195 IITPLSA 201  
|||||  
Db 192 IITPLSA 199

RESULT 18  
F87414  
phage SP01 DNA polymerase-related protein [imported] - Caulobacter crescentus  
C:Species: Caulobacter crescentus  
C:Date: 20-Apr-2001 #sequence\_revision 20-Apr-2001 #text\_change 20-Apr-2001  
C:Accession: F87414  
R:Nierman, W.C.; Feldblyum, T.V.; Paulsen, I.T.; Nelson, K.E.; Eisen, J.; Heidelberg, J  
n. J.; Ermolaeva, M.; White, O.; Salzberg, S.L.; Shapiro, L.; Venter, J.C.; Fraser, C.M  
B.; Laub, M.T.; DeBoy, R.T.; Dodson, R.J.; Durkin, A.S.; Gwinn, M.L.; Haft, D.H.; Kolo  
Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001  
A:Title: Complete Genome Sequence of Caulobacter crescentus.  
A:Reference number: A87249; MUID:21173696; PMID:11259647  
A:Accession: F87414  
A>Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-273 <STO>  
A:Cross-references: GB:AE005673; NID:g13422678; PIDN:AAK23314.1; GSPDB:GN00148  
C:Genetics:  
A:Gene: CCl333

Query Match 2.9%; Score 7; DB 2; Length 273;  
Best Local Similarity 100.0%; Pred. No. 51;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 100 EGPQADE 106  
|||||  
Db 123 EGPQADE 129

RESULT 19

F89796  
hypothetical protein SA0305 [imported] - Staphylococcus aureus (strain N315)  
C:Species: Staphylococcus aureus  
C>Date: 10-May-2001 #sequence\_revision 10-May-2001 #text\_change 22-Oct-2001  
C/Accession: F89796  
R:Kuroda, M.; Ohta, T.; Uchiyama, I.; Baba, T.; Yuzawa, H.; Kobayashi, I.; Cui, L.; Oguc  
ma, A.; Mizukani-Ui, Y.; Kobayashi, N.; Sawano, T.; Inoue, R.; Kaito, C.; Sekimizu, K.;  
C.; Shiba, T.; Hattori, M.; Ogasawara, N.; Hayashi, H.; Hiramatsu, K.  
Lancet 357, 1225-1240, 2001  
A:Title: Whole genome sequencing of methicillin-resistant Staphylococcus aureus.  
A:Reference number: A89758; MUID:21311952; PMID:11418146  
A/Accession: F89796  
A/Status: preliminary  
A:Note: the authors translated the codon CTG for residue 180 as Ser  
A:Gene: SA0305  
A:Residues: 1-286 <KUR>  
A/Cross-references: GB:BA000018; PID:G13700231; PIDN:BA841529.1; GSPDB:GN00149  
A:Experimental source: strain N315  
C/Genetics:  
A:Map position: 5  
A:Introns: 273/1

Query Match 2.9%; Score 7; DB 2; Length 286;  
Best Local Similarity 100.0%; Pred. No. 53;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 126 QPVIGIS 132  
Db 59 QPVIGIS 65  
|||||

RESULT 20  
D90460  
conserved hypothetical protein [imported] - Sulfolobus solfataricus  
C:Species: Sulfolobus solfataricus  
C/Date: 24-May-2001 #sequence\_revision 24-May-2001 #text\_change 24-May-2001  
C/Accession: D90460  
R:She, Q.; Singh, R.K.; Confalonieri, F.; Zivanovic, Y.; Allard, G.; Awayez, M.J.; Chan-  
Jong, I.; Jeffries, A.C.; Kozera, C.J.; Medina, N.; Peng, X.; Thi-Ngoc, H.P.; Redder, P.  
arrett, R.A.; Ragan, M.A.; Sersen, C.W.; Van der Oost, J.  
submitted to GenBank, April 2001  
A:Description: Sulfolobus solfataricus complete genome.  
A/Reference number: A99139  
A/Accession: D90460  
A/Status: preliminary  
A:Note: the authors translated the codon CTG for residue 180 as Ser  
A:Gene: SSO2829  
A:Residues: 1-291 <KUR>  
A/Cross-references: GB:AE006641; NID:G13816184; PIDN:AAK42939.1; GSPDB:GN00155  
C/Genetics:  
A:Map position: 5  
A:Introns: 273/1

Query Match 2.9%; Score 7; DB 2; Length 291;  
Best Local Similarity 100.0%; Pred. No. 53;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 158 VIIIAIG 164  
Db 50 VIIIAIG 56  
|||||

RESULT 21  
S57923  
SEC14 protein - yeast (Candida albicans)  
C:Species: Candida albicans  
C/Date: 13-Jan-1996 #sequence\_revision 01-Mar-1996 #text\_change 26-Feb-1998  
C/Accession: S57923; S72193  
R:Monteoliva, L.; Sanchez, M.; Pla, J.; Gil, C.; Nombela, C.  
submitted to the EMBL Data Library, September 1994  
A:Description: Characterisation of the Candida albicans SEC14 homolog gene.  
A/Reference number: S57923  
A/Accession: S57923  
A/Status: preliminary  
A:Note: the authors translated the codon CTG for residue 180 as Ser  
A:Gene: CESP:T15B7.5  
A:Residues: 1-301 <MON>  
A/Cross-references: EMBL:X81937

Query Match 2.9%; Score 7; DB 2; Length 301;  
Best Local Similarity 100.0%; Pred. No. 55;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 168 ILGYSYK 174  
Db 242 ILGYSYK 248  
|||||

RESULT 22  
T28999  
hypothetical protein ZC513.8 - Caenorhabditis elegans  
C:Species: Caenorhabditis elegans  
C/Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 15-Sep-2003  
C/Accession: T28999  
R:Wu, X.; Le, T.T.  
submitted to the EMBL Data Library, April 1996  
A:Description: The sequence of C. elegans cosmid ZC513.  
A/Reference number: Z20551  
A/Accession: T28999  
A/Status: preliminary; translated from GB/EMBL/DBJ  
A:Note: the authors translated the codon CTG for residue 180 as Ser  
A:Gene: CESP:ZC513.8  
A:Map position: 5  
A:Introns: 55/3; 229/1

Query Match 2.9%; Score 7; DB 2; Length 303;  
Best Local Similarity 100.0%; Pred. No. 55;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 235 GQAGTPG 241  
Db 274 GQAGTPG 280  
|||||

RESULT 23  
T32247  
hypothetical protein T15B7.5 - Caenorhabditis elegans  
C:Species: Caenorhabditis elegans  
C/Date: 29-Oct-1999 #sequence\_revision 29-Oct-1999 #text\_change 15-Sep-2003  
C/Accession: T32247  
R:Pauley, A.; Gattung, S.  
submitted to the EMBL Data Library, September 1997  
A:Description: The sequence of C. elegans cosmid T15B7.  
A/Reference number: Z21139  
A/Accession: T32247  
A/Status: preliminary; translated from GB/EMBL/DBJ  
A:Note: the authors translated the codon CTG for residue 180 as Ser  
A:Gene: CESP:T15B7.5  
A:Map position: 5  
A:Introns: 273/1

R;Monteoliva, L.; Sanchez, M.; Pla, J.; Gil, C.; Nombela, C.  
Yeast 12, 1097-1105, 1996  
A:Title: Cloning of Candida albicans SEC14 gene homologue coding for a putative essenti-  
A:Reference number: S72193; MUID:97051600; PMID:8896277  
A/Accession: S72193  
A/Molecule type: DNA  
A:Residues: 1-301 <MON>  
A/Cross-references: EMBL:X81937  
A:Note: the authors translated the codon CTG for residue 180 as Ser  
A:Gene: SEC14  
C/Superfamily: cellular retinaldehyde-binding protein; cellular retinaldehyde-binding p  
F:59-266/Domain: cellular retinaldehyde-binding protein homology <CRB>

Query Match 2.9%; Score 7; DB 2; Length 301;  
Best Local Similarity 100.0%; Pred. No. 55;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 168 ILGYSYK 174  
Db 242 ILGYSYK 248  
|||||

RESULT 22  
T28999  
hypothetical protein ZC513.8 - Caenorhabditis elegans  
C:Species: Caenorhabditis elegans  
C/Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 15-Sep-2003  
C/Accession: T28999  
R:Wu, X.; Le, T.T.  
submitted to the EMBL Data Library, April 1996  
A:Description: The sequence of C. elegans cosmid ZC513.  
A/Reference number: Z20551  
A/Accession: T28999  
A/Status: preliminary; translated from GB/EMBL/DBJ  
A:Note: the authors translated the codon CTG for residue 180 as Ser  
A:Gene: CESP:ZC513.8  
A:Map position: 5  
A:Introns: 55/3; 229/1

Query Match 2.9%; Score 7; DB 2; Length 303;  
Best Local Similarity 100.0%; Pred. No. 55;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 235 GQAGTPG 241  
Db 274 GQAGTPG 280  
|||||

RESULT 23  
T32247  
hypothetical protein T15B7.5 - Caenorhabditis elegans  
C:Species: Caenorhabditis elegans  
C/Date: 29-Oct-1999 #sequence\_revision 29-Oct-1999 #text\_change 15-Sep-2003  
C/Accession: T32247  
R:Pauley, A.; Gattung, S.  
submitted to the EMBL Data Library, September 1997  
A:Description: The sequence of C. elegans cosmid T15B7.  
A/Reference number: Z21139  
A/Accession: T32247  
A/Status: preliminary; translated from GB/EMBL/DBJ  
A:Note: the authors translated the codon CTG for residue 180 as Ser  
A:Gene: CESP:T15B7.5  
A:Map position: 5  
A:Introns: 273/1

```

Query Match          2.9%; Score 7; DB 2; Length 314;
Best Local Similarity 100.0%; Pred. No. 57;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 235 GQAGTPG 241
DB 251 GQAGTPG 257

RESULT 24
AD2612
iron-sulfur cluster binding protein [imported] - Agrobacterium tumefaciens (strain C58,
C:Species: Agrobacterium tumefaciens
C:Date: 11-Jan-2002 #sequence_revision 11-Jan-2002 #text_change 18-Nov-2002
C:Accession: AD2612
R:Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.B.; Chen, Y.; Woo, I.;
Karp, P.; Romero, P.; Grant, C.; Guenther, D.; Kutyavin, T.; Levy, R.; Li, M.; McClell
Science 294, 2317-2323, 2001
A:Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kamm,
star, E.W.
A:Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.
A:Reference number: AB2577; MUID:21608550; PMID:11743193
A:Accession: AD2612
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-326 <KUR>
A:Cross-references: GB:AE008698; PIDN:RAL41314.1; PID:gl7738625; GSPDB:GN00186
A:Experimental source: strain C58 (Dupont)
C:Genetics:
A:Gene: Atu0292
A:Map position: circular chromosome

Query Match          2.9%; Score 7; DB 2; Length 326;
Best Local Similarity 100.0%; Pred. No. 59;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 52 PDEDPRG 58
DB 33 PDEDPRG 39

RESULT 25
B97394
hypothetical protein AGR_C_502 [imported] - Agrobacterium tumefaciens (strain C58, Cerec
C:Species: Agrobacterium tumefaciens
C:Date: 30-Sep-2001 #sequence_revision 30-Sep-2001 #text_change 18-Nov-2002
C:Accession: B97394
R:Goodner, B.; Hinkle, G.; Gatrung, S.; Miller, N.; Blanchard, M.; Qurollo, B.; Goldman,
A.; Liu, F.; Wollam, C.; Allinger, M.; Doughty, D.; Scott, C.; Lappas, C.; Markelz, B.;
Science 294, 2323-2328, 2001
A:Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent Agrobacterium tum
A:Reference number: A97359; MUID:21608551; PMID:11743194
A:Accession: B97394
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-326 <KUR>
A:Cross-references: GB:AE007869; PIDN:AXK86107.1; PID:gl5155190; GSPDB:GN00169
C:Genetics:
A:Gene: AGR_C_502
A:Map position: circular chromosome

Query Match          2.9%; Score 7; DB 2; Length 326;
Best Local Similarity 100.0%; Pred. No. 59;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 52 PDEDPRG 58
DB 33 PDEDPRG 39

RESULT 26

```

```

AH1994
hypothetical protein all1509 [imported] - Nostoc sp. (strain PCC 7120)
C:Species: Nostoc sp. PCC 7120
A:Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120
C:Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 09-Dec-2002
C:Accession: AH1994
R:Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriguchi
Nakazaki, N.; Shimo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata,
DNA Res. 8, 205-213, 2001
A:Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Al
A:Reference number: AB1807; MUID:21595285; PMID:11759840
A:Accession: AH1994
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-332 <KUR>
A:Cross-references: GB:BA000019; PIDN:BAW77875.1; PID:gl7135330; GSPDB:GN00179
A:Experimental source: strain PCC 7120
C:Genetics:
A:Gene: all1509

Query Match          2.9%; Score 7; DB 2; Length 332;
Best Local Similarity 100.0%; Pred. No. 60;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 161 IAIGAGI 167
DB 221 IAIGAGI 227

RESULT 27
F64080
glycerol-3-phosphate dehydrogenase (NAD) (EC 1.1.1.18) - Haemophilus influenzae (strain
C:Species: Haemophilus influenzae
C:Date: 18-Aug-1995 #sequence_revision 18-Aug-1995 #text_change 27-Oct-2003
C:Accession: F64080
R:Fleischmann, R.D.; Adams, M.D.; White, O.; Clayton, R.A.; Kirkness, B.F.; Kerlavage,
; Gocayne, J.D.; Scott, J.; Shirley, R.; Liu, L.I.; Glodek, A.; Kelley, J.M.; Weidman,
D.M.; Brandon, R.C.; Fine, L.D.; Fritchman, J.L.; Fuhrmann, J.L.; Geoghegan, N.S.M.
Science 269, 496-512, 1995
A:Authors: Gnehm, C.L.; McDonald, L.A.; Small, K.V.; Fraser, C.M.; Smith, H.O.; Venter,
A:Title: Whole-genome random sequencing and assembly of Haemophilus influenzae Rd.
A:Reference number: A64000; MUID:95350630; PMID:7542800
A:Accession: F64080
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-335 <TGR>
A:Cross-references: GB:U32743; GB:L42023; NID:gl573597; PIDN:AAC22264.1; PID:gl573598;
C:Superfamily: glycerol-3-phosphate dehydrogenase (NAD)
C:Keywords: oxidoreductase

Query Match          2.9%; Score 7; DB 2; Length 335;
Best Local Similarity 100.0%; Pred. No. 60;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 161 IAIGAGI 167
DB 198 IAIGAGI 204

RESULT 28
D64397
hypothetical protein MJ0780 - Methanococcus jannaschii
C:Species: Methanococcus jannaschii
C:Date: 13-Sep-1996 #sequence_revision 13-Sep-1996 #text_change 21-Jul-2000
C:Accession: D64397
R:Bult, C.J.; White, O.; Olsen, G.J.; Zhou, L.; Fleischmann, R.D.; Sutton, G.G.; Blake,
; Reich, C.I.; Overbeek, R.; Kirkness, E.F.; Weinstock, K.G.; Merrick, J.M.; Glodek, A.
son, J.D.; Sadow, P.W.; Hanna, M.C.; Cotton, M.D.; Roberts, K.M.; Hurst, M.A.
Science 273, 1058-1073, 1996
A:Authors: Kaine, B.P.; Borodovsky, M.; Klenk, H.P.; Fraser, C.M.; Smith, H.O.; Woese,
A:Title: Complete genome sequence of the methanogenic archaeon, Methanococcus jannaschi
A:Reference number: A64300; MUID:96337999; PMID:8688087
A:Accession: D64397

```

```
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-335 <BUL>
A;Cross-references: GB:U67522; GB:L77117; NID:G2826315; PIDN:AAB98779.1; PID:gl499600; T
C;Genetics:
A;Map position: REV701539-700532

Query Match      2.9%; Score 7; DB 2; Length 335;
Best Local Similarity 100.0%; Pred. No. 60;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 163 IGAGIIL 169
    |||||
Db 137 IGAGIIL 143

RESULT 29
AB2129
iron(III) diclitate transport system permease protein all2585 [imported] - Nostoc sp. (s
C;Species: Nostoc sp. PCC 7120
A;Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120
C;Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 09-Dec-2002
C;Accession: AB2129
R;Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasanoto, S.; Watanabe, A.; Iriiguchi
Nakazaki, N.; Shimo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata, S
DNA Res. 8, 205-213, 2001
A;Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Ana
A;Reference number: AB1807; MUID:21595285; PMID:11759840
A;Accession: AB2129
A;Status: preliminary
A;Molecule type: DNA
A;Cross-references: GB:BA000019; PIDN:BAE74284.1; PID:gl7131678; GSPDB:GN00179
C;Genetics:
A;Gene: all2585
C;Superfamily: vitamin B12 transport protein btuC

Query Match      2.9%; Score 7; DB 2; Length 346;
Best Local Similarity 100.0%; Pred. No. 62;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 196 TLPLSAP 202
    |||||
Db 133 TLPLSAP 139

RESULT 30
AB6590
hypothetical protein CPj0795 [imported] - Chlamydomophila pneumoniae (strain J138)
C;Species: Chlamydomophila pneumoniae, Chlamydia pneumoniae
C;Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 02-Mar-2001
C;Accession: AB6590
R;Shirai, M.; Hirakawa, H.; Kimoto, M.; Tabuchi, M.; Kishi, F.; Ouchi, K.; Shiba, T.; Ie
Nucleic Acids Res. 28, 2311-2314, 2000
A;Title: Comparison of whole genome sequences of chlamydia pneumoniae J138.
A;Reference number: AB6491; MUID:20330349; PMID:10871362
A;Accession: AB6590
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-336 <STO>
A;Cross-references: GB:BA000008; NID:G8979168; PIDN:BA999003.1; GSPDB:GN00142
A;Experimental source: strain J138
C;Genetics:
A;Gene: CPj0795

Query Match      2.9%; Score 7; DB 2; Length 356;
Best Local Similarity 100.0%; Pred. No. 64;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 143 KDLGTLG 149
    |||||
Db 2 KDLGTLG 8

RESULT 33
B81505
hypothetical protein CP1074 [imported] - Chlamydomophila pneumoniae (strain AR39)
C;Species: Chlamydomophila pneumoniae, Chlamydia pneumoniae
```

## RESULT 31

```
D72033
hypothetical protein CP1076 [imported] - Chlamydomophila pneumoniae (strains CWL029 and A
C;Species: Chlamydomophila pneumoniae, Chlamydia pneumoniae
C;Date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change 11-May-2000
C;Accession: D72033; D81505
R;Kalman, S.; Mitchell, W.; Marathe, R.; Lammel, C.; Fan, J.; Olinger, L.; Grimwood, J.
Nature Genet. 21, 385-389, 1999
A;Title: Comparative genomes of Chlamydia pneumoniae and C. trachomatis.
A;Reference number: A72000; MUID:99206606; PMID:10192388
A;Accession: D72033
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-356 <ARN>
A;Cross-references: GB:AE001661; GB:AE001363; NID:G4377104; PIDN:AAD18933.1; PID:G43771
A;Experimental source: strain CWL029
R;Read, T.D.; Brunham, R.C.; Shen, C.; Gill, S.R.; Heidelberg, J.F.; White, O.; Hickey,
C.; Dodson, R.; Gwinn, M.; Nelson, W.; DeBoy, R.; Kolonay, J.; McClarty, G.; Salzberg
Nucleic Acids Res. 28, 1397-1406, 2000
A;Title: Genome sequences of Chlamydia trachomatis MoPn and Chlamydia pneumoniae AR39.
A;Reference number: A81500; MUID:20150255; PMID:10684935
A;Accession: D81505
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-356 <REA>
A;Cross-references: GB:AE002264; GB:AE002161; NID:G7189984; PIDN:AAF38848.1; PID:G71899
A;Experimental source: strain AR39, HL cells
C;Genetics:
A;Gene: CPn0795; CP1076
```

```
Query Match      2.9%; Score 7; DB 2; Length 356;
Best Local Similarity 100.0%; Pred. No. 64;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 143 KDLGTLG 149
    |||||
Db 2 KDLGTLG 8
```

## RESULT 32

```
F72033
hypothetical protein - Chlamydomophila pneumoniae (strain CWL029)
C;Species: Chlamydomophila pneumoniae, Chlamydia pneumoniae
C;Date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change 05-May-2000
C;Accession: F72033
R;Kalman, S.; Mitchell, W.; Marathe, R.; Lammel, C.; Fan, J.; Olinger, L.; Grimwood, J.
Nature Genet. 21, 385-389, 1999
A;Title: Comparative genomes of Chlamydia pneumoniae and C. trachomatis.
A;Reference number: A72000; MUID:99206606; PMID:10192388
A;Accession: F72033
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-365 <ARN>
A;Cross-references: GB:AE001661; GB:AE001363; NID:G4377104; PIDN:AAD18935.1; PID:G43771
A;Experimental source: strain CWL029
C;Genetics:
A;Gene: CPn0797
```

```
Query Match      2.9%; Score 7; DB 2; Length 365;
Best Local Similarity 100.0%; Pred. No. 65;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 143 KDLGTLG 149
    |||||
Db 142 KDLGTLG 148
```

## RESULT 33

```
B81505
hypothetical protein CP1074 [imported] - Chlamydomophila pneumoniae (strain AR39)
C;Species: Chlamydomophila pneumoniae, Chlamydia pneumoniae
```

C;Date: 31-Mar-2000 #sequence\_revision 31-Mar-2000 #text\_change 11-May-2000  
 C;Accession: B81505  
 R;Read, T.D.; Brunham, R.C.; Shen, C.; Gill, S.R.; Heidelberg, J.F.; White, O.; Hickey,  
 C.; Dodson, R.; Gwinn, M.; Nelson, W.; DeBoy, R.; Kolonay, J.; McClarty, G.; Salzberg,  
 Nucleic Acids Res. 28, 1397-1406, 2000  
 A;Title: Genome sequences of Chlamydia trachomatis MoPn and Chlamydia pneumoniae AR39.  
 A;Reference number: AB1500; MUID:20150255; PMID:10684935  
 A;Accession: B81505  
 A;Status: preliminary  
 A;Molecule type: DNA  
 A;Residues: 1-365 <REA>  
 A;Cross-references: GB:AE002264; GB:AE002161; NID:g7189984; PIDN:AAF38846.1; PID:g718998  
 A;Experimental source: strain AR39, HL cells  
 C;Genetics:  
 A;Gene: CPl074

Query Match 2.9%; Score 7; DB 2; Length 365;  
 Best Local Similarity 100.0%; Pred. No. 65;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 143 KDLGTLG 149  
 |||||  
 DB 142 KDLGTLG 148

## RESULT 34

C86590  
 hypochloral protein CPJ0797 [imported] - Chlamydothila pneumoniae (strain J138)  
 C;Species: Chlamydothila pneumoniae, Chlamydia pneumoniae  
 C;Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 02-Mar-2001  
 C;Accession: C86590  
 R;Shirai, M.; Hirakawa, H.; Kimoto, M.; Tabuchi, M.; Kishi, F.; Ouchi, K.; Shiba, T.; Ie  
 Nucleic Acids Res. 28, 2311-2314, 2000  
 A;Title: Comparison of whole genome sequences of chlamydia pneumoniae J138.  
 A;Reference number: AB6491; MUID:20330349; PMID:10871362  
 A;Accession: C86590  
 A;Status: preliminary  
 A;Molecule type: DNA  
 A;Residues: 1-365 <STO>  
 A;Cross-references: GB:BAC00008; NID:g8979171; PIDN:BAA99005.1; GSPDB:GN00142  
 A;Experimental source: strain J138  
 C;Genetics:  
 A;Gene: CPJ0797

Query Match 2.9%; Score 7; DB 2; Length 365;  
 Best Local Similarity 100.0%; Pred. No. 65;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 143 KDLGTLG 149  
 |||||  
 DB 142 KDLGTLG 148

## RESULT 35

S18560  
 xylR protein - Lactobacillus pentosus  
 C;Species: Lactobacillus pentosus  
 C;Date: 22-Nov-1993 #sequence\_revision 26-May-1995 #text\_change 15-Oct-1999  
 C;Accession: S18560  
 R;Lokman, B.C.; van Santen, P.; Verdoes, J.C.; Kruese, J.; Leer, R.J.; Posno, M.; Pouwel  
 Mol. Gen. Genet. 230, 161-169, 1991  
 A;Title: Organization and characterization of three genes involved in D-xylose catabolism  
 A;Reference number: S18560; MUID:92079891; PMID:1660563  
 A;Accession: S18560  
 A;Status: preliminary  
 A;Molecule type: DNA  
 A;Residues: 1-388 <LOK>  
 A;Cross-references: EMBL:M57384; NID:gl49604; PIDN:AAA25257.1; PID:gl49605  
 A;Note: the authors translated the initiation codon GTG for residue 1 as Val  
 C;Genetics:  
 A;Gene: xylR  
 A;Start codon: GTG  
 C;Superfamily: xylose repressor; glucose kinase homology

C;Keywords: DNA binding; transcription regulation  
 F144-267/Domain: glucose kinase homology <GKH>

Query Match 2.9%; Score 7; DB 2; Length 388;  
 Best Local Similarity 100.0%; Pred. No. 69;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 163 IGAGIIL 169  
 |||||  
 DB 219 IGAGIIL 225

## RESULT 36

B83252  
 probable amino acid aminotransferase PA3139 [imported] - Pseudomonas aeruginosa (strain  
 C;Species: Pseudomonas aeruginosa  
 C;Date: 15-Sep-2000 #sequence\_revision 15-Sep-2000 #text\_change 31-Dec-2000  
 A;Accession: B83252  
 R;Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warrenner, P.; Hickey, M.J.; B  
 adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; Lim  
 ; Lory, S.; Olson, M.V.  
 Nature 405, 959-964, 2000  
 A;Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic path  
 A;Reference number: AB2950; MUID:20437337; PMID:10984043  
 A;Accession: B83252  
 A;Status: preliminary  
 A;Molecule type: DNA  
 A;Residues: 1-398 <STO>  
 A;Cross-references: GB:AE004738; GB:AE004091; NID:g9949252; PIDN:AAG06527.1; GSPDB:GN00  
 A;Experimental source: strain PA01  
 C;Genetics:  
 A;Gene: PA3139  
 C;Superfamily: aspartate aminotransferase

Query Match 2.9%; Score 7; DB 2; Length 398;  
 Best Local Similarity 100.0%; Pred. No. 70;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 114 NALPARS 120  
 |||||  
 DB 169 NALPARS 175

## RESULT 37

T43406  
 cullin-3 - fission yeast (Schizosaccharomyces pombe)  
 C;Species: Schizosaccharomyces pombe  
 C;Date: 21-Jan-2000 #sequence\_revision 21-Jan-2000 #text\_change 21-Jan-2000  
 C;Accession: T43406  
 R;Kominami, K.; Toda, T.  
 submitted to the EMBL Data Library, August 1998  
 A;Description: Fcu3 (S. pombe cullin-3).  
 A;Reference number: Z22490  
 A;Accession: T43406  
 A;Status: preliminary; translated from GB/EMBL/DBJ  
 A;Molecule type: DNA  
 A;Residues: 1-421 <KOM>  
 A;Cross-references: EMBL:AB017028; PIDN:BAA32519.1  
 C;Genetics:  
 A;Gene: Fcu3  
 C;Function:  
 A;Description: involved in the distinct stress-response pathway  
 A;Note: not a component of SCFPop1,2 like cullin-1

Query Match 2.9%; Score 7; DB 2; Length 421;  
 Best Local Similarity 100.0%; Pred. No. 74;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 82 ETTSQAL 88  
 |||||  
 DB 218 ETTSQAL 224

```
RESULT 38
AB2740
A:Accession: T06598
A:Molecule type: DNA
A:Residues: 1-448 <QUE>
A:Cross-references: EMBL:AL049659; GSPDB:GN00061; ATSP:T29H11.40
C:Date: 11-Jan-2002 #sequence_revision 11-Jan-2002 #text_change 18-Nov-2002
C:Accession: AB2740
R:Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo, I.;
erge, G.; Gillet, W.; Grant, C.; Guenther, D.; Kutyavin, T.; Levy, R.; Li, M.; McClell
; Karp, P.; Romero, P.; Zhang, S.
Science 294, 2317-2323, 2001
A:Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kamm,
ster, E.W.
A:Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.
A:Reference number: AB2577; MUID:21608550; PMID:11743193
A:Accession: AB2740
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-448 <KUR>
A:Cross-references: GB:AE008688; PIDN:AAL42336.1; PID:gl7739740; GSPDB:GN00186
A:Experimental source: strain C58 (Dupont)
C:Genetics:
A:Gene: accC
A:Map position: circular chromosome
C:Superfamily: biotin carboxylase; biotin carboxylase homology

Query Match      2.9%; Score 7; DB 2; Length 448;
Best Local Similarity 100.0%; Pred. No. 78;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 118 ARSEAAA 124
      |||||
DB 183 ARSEAAA 189

RESULT 39
H97520
A:Accession: H97520
A:Molecule type: DNA
A:Residues: 1-448 <KUR>
A:Cross-references: GB:AE007869; PIDN:AAK87121.1; PID:gl5156385; GSPDB:GN00169
C:Genetics:
A:Gene: AGS C 2451
A:Map position: circular chromosome
C:Superfamily: biotin carboxylase; biotin carboxylase homology

Query Match      2.9%; Score 7; DB 2; Length 448;
Best Local Similarity 100.0%; Pred. No. 78;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 118 ARSEAAA 124
      |||||
DB 183 ARSEAAA 189

RESULT 40
T06698
A:Accession: T06698
A:Molecule type: DNA
A:Residues: 1-448 <QUE>
A:Cross-references: EMBL:AL049659; GSPDB:GN00061; ATSP:T29H11.40
C:Date: 11-Jan-2002 #sequence_revision 11-Jan-2002 #text_change 18-Nov-2002
C:Accession: AB2740
R:Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo, I.;
erge, G.; Gillet, W.; Grant, C.; Guenther, D.; Kutyavin, T.; Levy, R.; Li, M.; McClell
; Karp, P.; Romero, P.; Zhang, S.
Science 294, 2317-2323, 2001
A:Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kamm,
ster, E.W.
A:Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.
A:Reference number: AB2577; MUID:21608550; PMID:11743193
A:Accession: AB2740
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-448 <KUR>
A:Cross-references: GB:AE008688; PIDN:AAL42336.1; PID:gl7739740; GSPDB:GN00186
A:Experimental source: strain C58 (Dupont)
C:Genetics:
A:Gene: accC
A:Map position: circular chromosome
C:Superfamily: biotin carboxylase; biotin carboxylase homology

Query Match      2.9%; Score 7; DB 2; Length 448;
Best Local Similarity 100.0%; Pred. No. 78;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 118 ARSEAAA 124
      |||||
DB 183 ARSEAAA 189

RESULT 41
AD0799
A:Accession: AD0799
A:Molecule type: DNA
A:Residues: 1-463 <PAR>
A:Cross-references: GB:AL513382; PIDN:CAD07574.1; PID:gl6503566; GSPDB:GN00176
C:Genetics:
A:Gene: STY2572

Query Match      2.9%; Score 7; DB 2; Length 463;
Best Local Similarity 100.0%; Pred. No. 80;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 146 GTLGYVL 152
      |||||
DB 188 GTLGYVL 194

RESULT 42
AS3918
A:Accession: AS3918
A:Molecule type: mRNA
A:Residues: 1-483 <KRI>
A:Cross-references: GB:U10422; NID:G533504; PIDN:AAA61639.1; PID:G533505
C:Keywords: glycosidase; hydrolase; polysaccharide degradation

Query Match      2.9%; Score 7; DB 2; Length 483;
Best Local Similarity 100.0%; Pred. No. 83;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 38 APVSGAG 44
      |||||
```

```
A:Accession: T06598
A:Molecule type: DNA
A:Residues: 1-448 <QUE>
A:Cross-references: EMBL:AL049659; GSPDB:GN00061; ATSP:T29H11.40
C:Date: 11-Jan-2002 #sequence_revision 11-Jan-2002 #text_change 18-Nov-2002
C:Accession: AB2740
R:Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo, I.;
erge, G.; Gillet, W.; Grant, C.; Guenther, D.; Kutyavin, T.; Levy, R.; Li, M.; McClell
; Karp, P.; Romero, P.; Zhang, S.
Science 294, 2317-2323, 2001
A:Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kamm,
ster, E.W.
A:Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.
A:Reference number: AB2577; MUID:21608550; PMID:11743193
A:Accession: AB2740
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-448 <KUR>
A:Cross-references: GB:AE008688; PIDN:AAL42336.1; PID:gl7739740; GSPDB:GN00186
A:Experimental source: strain C58 (Dupont)
C:Genetics:
A:Gene: accC
A:Map position: circular chromosome
C:Superfamily: biotin carboxylase; biotin carboxylase homology

Query Match      2.9%; Score 7; DB 2; Length 448;
Best Local Similarity 100.0%; Pred. No. 78;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 33 SGLASAP 39
      |||||
DB 188 SGLASAP 194

RESULT 41
AD0799
A:Accession: AD0799
A:Molecule type: DNA
A:Residues: 1-463 <PAR>
A:Cross-references: GB:AL513382; PIDN:CAD07574.1; PID:gl6503566; GSPDB:GN00176
C:Genetics:
A:Gene: STY2572

Query Match      2.9%; Score 7; DB 2; Length 463;
Best Local Similarity 100.0%; Pred. No. 80;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 146 GTLGYVL 152
      |||||
DB 188 GTLGYVL 194

RESULT 42
AS3918
A:Accession: AS3918
A:Molecule type: mRNA
A:Residues: 1-483 <KRI>
A:Cross-references: GB:U10422; NID:G533504; PIDN:AAA61639.1; PID:G533505
C:Keywords: glycosidase; hydrolase; polysaccharide degradation

Query Match      2.9%; Score 7; DB 2; Length 483;
Best Local Similarity 100.0%; Pred. No. 83;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 38 APVSGAG 44
      |||||
```

281 APVSGAG 287

RESULT 43

AB0397  
Multidrug resistance protein B [imported] - Yersinia pestis (strain CO92)  
C/Species: Yersinia pestis  
C/Date: 02-Nov-2001 #sequence\_revision 02-Nov-2001 #text\_change 17-May-2002  
C/Accession: AB0397  
R:/Parkhill, J.; Wren, B.W.; Thomson, N.R.; Titball, R.W.; Holden, M.T.G.; Prentice, M.B.;  
deleno-Parraga, A.M.; Chillingworth, T.; Cronin, A.; Davies, R.M.; Davis, P.; Dougan, G.;  
Gil, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.; Whitehead, S.; Barrall,  
Nature 413, 523-527, 2001  
A/TITLE: Genome sequence of Yersinia pestis, the causative agent of plague.  
A/Reference number: AB0001; MUID:21470413; PMID:11586360  
A/Accession: AB0397  
A/Status: Preliminary  
A/Molecule type: DNA  
A/Residues: 1-511 <KUR>  
A/Cross-references: GB:AL590842; PIDN:CAC92502.1; PID:gl5981202; GSFDB:GN00175  
C/Genetics:  
A/Gene: emrB  
C/Superfamily: lincomycin-resistance protein lmrB

```
Query Match          2.9%; Score 7; DB 2; Length 511;
Best Local Similarity 100.0%; Pred.No. 87;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

34 GLASAPV 40  
306 GLASAPV 312

RESULT 44  
G65048  
multidrug resistance protein B - Escherichia coli (strain K-12)  
C:Species: Escherichia coli  
C:Date: 12-Sep-1997 #sequence revision 17-Sep-1997 #text\_change 01-Mar-2002  
C:Accession: G65048; JCI345; S57558  
R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Co  
A.; Rose, D.J.; Mau, B.; Shao, Y.  
Science 277, 1453-1462, 1997  
A:Title: The complete genome sequence of Escherichia coli K-12.  
A:Reference number: A64720; MUID:97436617; PMID:9278503  
A:Accession: G65048  
A:Status: nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-512 <BLAST>  
A:Cross-references: GB:AE000353; GB:U00096; NID:G1789037; PIDN:AAC75733.1; PID:G1789042;  
A:Experimental source: strain K-12, substrain MG1655

Proc. Natl. Acad. Sci. U.S.A. 89, 8938-8942, 1992

A: Title: emr, an *Escherichia coli* locus for multidrug resistance.

A: Reference number: JCI1344; MUID: 93028382; PMID: 1409590

A: Accession: JCI1345

A: Molecule type: DNA

A: Residues: 1-324, 'A', 326-500, 'A', 502-512 <LON>

A: Cross-references: GB: W86657; NID: g145834; PIDN: AAA23725.1; PID: g145836

C: Comment: This protein is resistant to carbonyl cyanide m-chlorophenylhydrazone, nalidixic acid, and tetracycline.

C: Genetics:

A: Gene: emrB

A: Map position: 57.5 min

C: Superfamily: lincomycin-resistance protein lmrB

C: Keywords: transmembrane protein

Query Match 2.9%; Score 7; DB 2; Length 512;  
Best Local Similarity 100.0%; Pred. No. 87;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

34 GLASAPV 40  
307 GLASAPV 313

RESULT 45  
D91072

multidrug resistance membrane translocase [imported] - Escherichia coli (strain O157:H7)  
 C:Species: Escherichia coli  
 C:Date: 18-Jul-2001 #sequence\_revision: 18-Jul-2001 #text\_change: 17-May-2002  
 C:Accession: D91072  
 R:Hayaashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.C.  
 gasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.  
 DNA Res. 8, 11-22, 2001  
 A:Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and gen  
 A:Reference numbers: A99629; PMID:21156231; PMID:11258796

A:Accession:U01312  
A:Accession:U01312  
A:Material:Relaxin DNA  
A:Material:Relaxin DNA  
A:Residues:1-512 <HAV>  
A:Cross-references: PIDN:BAE36971.1;  
A:Experimental source: strain O157:H7, substrain 1  
C:Genetics:  
C:Genetics:  
C:Gene: ECa3548  
C:Superfamily: lincomycin-resistance protein lmrB

```
Query Match      2.9%; Score 7; DB 2; Length 512;
Best Local Similarity 100.0%; Pred. No. 87;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

34 GLASAPV 40  
307 GLASAPV 313

RESULT 46

hypothetical protein emrB [imported] - Escherichia coli (strain O157:H7, substrain EDL33)  
C:Species: Escherichia coli  
C:Date: 16-Feb-2001 #sequence\_revision 16-Feb-2001 #text\_change 17-May-2002  
C:Accession: G85916  
R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew  
Miller, L.; Grobeck, E.J.; Davis, N.W.; Lim, A.; Dimaienta, B.; Potamousis, K.; Apodaca  
Nature 409, 523-533, 2001  
A:Title: Genome number: A85480; MUID:21074933; PMID:11206551  
A:Reference number: A85480; MUID:21074933; PMID:11206551

A/Accession: G85916  
A/Status: preliminary  
A/Molecule type: DNA  
A/Residues: 1-512 <SPO>  
A/Cross-references: GB:AE051174; NID:g12517127; PIN:AA657795.1; GSPDB:GN00145; UWGP:Z3  
A/Experimental source: strain O157:H7, substrain EDL933  
C/Genetics:  
A/Gene: emtB

**УТВЕРЖДАЮ:**

Query Match 2.9%; Score 7; DB 2; Length 512;  
Best Local Similarity 100.0%; Pred. No. 87;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

34 GLASAPV 40  
307 GLASAPV 313

RESULT 47  
AC0812

nucleating resistance protein B [imported] - Salmonella enterica subsp. enterica serovar  
 Anatum; Salmonella enterica subsp. enterica serovar Typhi  
 CNG01; Salmonella has also been called Salmonella typhi  
 CNG01; 09-Nov-2001 #sequence\_revision 09-Nov-2001 #text\_change 18-Nov-2002  
 R;Accession: AG0842  
 R;Pack:hill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Churcher  
 th, T.; Connerton, P.; Cronin, A.; Davis, P.; Davies, R.M.; Dowd, L.; White, N.; Farrar  
 , S.; Moule, S.; O'Gaora, P.

*Authors:* Barry C. Quail, M. Rutherford, K. J. Simmonds, M. Skelton, J. J. Stevens, K. K. S. Moule, S. J. O. Guala, F. Nature 413, 848-852, 2001

A;Title: Complete genome sequence of a multiple drug resistant Salmonella enterica serov  
A;Reference number: AB0502; MUID:21334947; PMID:11677608  
A;Accession: AG0842  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-512 <PAR>  
A;Cross-references: GB:AL513382; PIDN:CAD05926.1; PID:gi6503897; GSPDB:GN00176  
C;Genetics:  
A;Gene: SRY2941  
C;Superfamily: lincomycin-resistance protein lmrB

Query Match 2.9%; Score 7; DB 2; Length 512;  
Best Local Similarity 100.0%; Pred. No. 87;  
Matches 7; Conservative 0; Mismatches 0; Gaps 0;  
Indels 0;

Qy 34 GLASAPV 40  
|||  
Db 307 GLASAPV 313

RESULT 48  
AI3295  
adenylosuccinate synthase (EC 6.3.4.4) [imported] - Brucella melitensis (strain 16M)  
C;Species: Brucella melitensis  
C;Date: 01-Feb-2002 #sequence\_revision 01-Feb-2002 #text\_change 17-May-2002  
C;Accession: AI3295  
R;DelVecchio, V.G.; Kaprat, V.; Redkar, R.J.; Patra, G.; Mujer, C.; Los, T.; Ivanova,  
M.; Mazur, M.; Goldsman, E.; Selkov, E.; Elzer, P.H.; Hagius, S.; O'Callaghan, D.; Letess  
Proc. Natl. Acad. Sci. U.S.A. 99, 443-448, 2002  
A;Title: The genome sequence of the facultative intracellular pathogen Brucella melitensis  
A;Reference number: AD3252; PMID:11756688  
A;Accession: AI3295  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-520 <KUR>  
A;Cross-references: GB:AE008917; PIDN:AAL51532.1; PID:gi7982249; GSPDB:GN00190  
A;Experimental source: strain 16M  
C;Genetics:  
A;Gene: BMEI0351  
A;Map position: 1  
C;Superfamily: adenylosuccinate synthase  
C;Keywords: ligase

Query Match 2.9%; Score 7; DB 2; Length 520;  
Best Local Similarity 100.0%; Pred. No. 89;  
Matches 7; Conservative 0; Mismatches 0; Gaps 0;  
Indels 0;

Qy 149 GYVLGIT 155  
|||  
Db 350 GYVLGIT 356

RESULT 49  
WHYU4  
tyrosine 3-monooxygenase (EC 1.14.16.2), splice form 4 - human  
N;Alternate names: tyrosine 3-hydroxylase  
N;Contains: tyrosine 3-monooxygenase, splice form 1; tyrosine 3-monooxygenase, splice fo  
C;Species: Homo sapiens (man)  
C;Date: 31-Mar-1989 #sequence\_revision 31-Mar-1989 #text\_change 16-Jun-2000  
C;Accession: A30002; A25825; A60201; JE0012; JE0013; JE0014; A27791; B27791; C27791; PNC  
R;Nagatsu, T.  
submitted to GenBank, December 1987  
A;Reference number: A94509  
A;Accession: A30002  
A;Molecule type: mRNA  
A;Residues: 1-528 <NAG1>  
A;Cross-references: GB:M17589; NID:g339680; PIDN:AAA61179.1; PID:g339681  
R;Kobayashi, K.; Ichinose, H.; Kishi, F.; Nakazawa, A.; Kurosawa, Y.; Fujita  
Biochem. Biophys. Res. Commun. 146, 971-975, 1987  
A;Title: Isolation of a novel cDNA clone for human tyrosine hydroxylase: alternative RNA  
A;Reference number: A90136; MUID:87298614; PMID:2887169  
A;Accession: A26825  
A;Molecule type: mRNA

A;Residues: 1-94 <NAG2>  
A;Cross-references: GB:M17589; NID:g339680; PIDN:AAA61179.1; PID:g339681  
R;Le Bourdellies, B.; Boularand, S.; Soni, C.; Horellou, P.; Dumas, S.; Grima, B.; Mallie  
J. Neurochem. 50, 988-991, 1988  
A;Title: Analysis of the 5' region of the human tyrosine hydroxylase gene: combinatoria  
A;Reference number: A60201; MUID:88117543; PMID:2892893  
A;Accession: A60201  
A;Molecule type: mRNA  
A;Residues: 1-65 <LEB>  
A;Cross-references: GB:M24790; NID:g556223; PIDN:AAA61174.1; PID:g556224  
R;Kobayashi, K.; Kaneda, N.; Ichinose, H.; Kishi, F.; Nakazawa, A.; Kurosawa, Y.; Fujit  
J. Biochem. 103, 907-912, 1988  
A;Title: Structure of the human tyrosine hydroxylase gene: alternative splicing from a  
A;Reference number: JE0012; MUID:89008200; PMID:2302075  
A;Accession: JE0012  
A;Molecule type: DNA  
A;Residues: 1-30,62-135 <KOB1>  
A;Cross-references: GB:D00269; NID:g220099; PIDN:BAA25094.1; PID:g2951764  
A;Experimental source: splice form 1  
A;Note: this splice form is produced by an alternative donor site within exon 1  
A;Accession: JE0013  
A;Molecule type: DNA  
A;Residues: 1-34,62-135 <KOB2>  
A;Cross-references: GB:D00269; NID:g220099; PIDN:BAA25097.1; PID:g2951767  
A;Experimental source: splice form 2  
A;Accession: JE0014  
A;Molecule type: DNA  
A;Residues: 1-30,35-135 <KOB3>  
A;Cross-references: GB:D00269; NID:g220099; PIDN:BAA25095.1; PID:g2951765  
A;Experimental source: splice form 3  
A;Note: this splice form is produced by an alternative donor site within exon 1  
R;Grima, B.; Lamouroux, A.; Boni, C.; Julien, J.F.; Javoy-Agid, F.; Mallet, J.  
Nature 326, 707-711, 1987  
A;Title: A single human gene encoding multiple tyrosine hydroxylases with different pre  
A;Reference number: A93393; MUID:87173064; PMID:2882428  
A;Accession: A27791  
A;Molecule type: mRNA  
A;Residues: 1-30,62-528 <GR1>  
A;Cross-references: GB:X05290; NID:g32501; PIDN:CRA28908.1; PID:g32502  
A;Experimental source: splice form 1  
A;Note: this splice form is produced by an alternative donor site within exon 1  
A;Accession: B27791  
A;Molecule type: mRNA  
A;Residues: 1-34,62-528 <GR2>  
A;Cross-references: GB:X05290; NID:g32501  
A;Experimental source: splice form 2  
A;Accession: C27791  
A;Molecule type: mRNA  
A;Residues: 30,35-528 <GR13>  
A;Cross-references: GB:X05290; NID:g32501  
A;Experimental source: splice form 3  
A;Note: this isozyne is produced by use of an alternative donor site within exon 1  
R;Ichinose, H.; Ohye, T.; Fujita, K.; Yoshida, M.; Ueda, S.; Nagatsu, T.  
Biochem. Biophys. Res. Commun. 195, 158-165, 1993  
A;Title: Increased heterogeneity of tyrosine hydroxylase in humans.  
A;Reference number: PNC0575; MUID:93371399; PMID:7689834  
A;Accession: PNC0575  
A;Status: translation not shown  
A;Molecule type: DNA  
A;Residues: 19-30 <ICH1>  
A;Accession: PNC0582  
A;Status: translation not shown  
A;Molecule type: DNA  
A;Residues: 35-61 <ICH2>  
A;Accession: PNC0588  
A;Status: translation not shown  
A;Molecule type: DNA  
A;Residues: 62-105 <ICH3>  
R;O'Malley, K.L.; Anhalt, M.J.; Martin, B.M.; Kelsoe, J.R.; Winfield, S.L.; Ginns, E.I.  
Biochemistry 26, 2910-2914, 1987  
A;Title: Isolation and characterization of the human tyrosine hydroxylase gene: identif  
A;Reference number: I52396; MUID:88107612; PMID:2892528  
A;Accession: I52396



A;Status: translated from GB/EMBL/DBJ  
A;Molecule type: DNA  
A;Residues: 1-61 <DNA>  
A;Cross-references: GB:M18116; NID:G339633; PIDN:AAA77649.1; PID:G1004335  
R; Kobayashi, K.; Kaneda, N.; Ichinose, H.; Kishi, F.; Nakazawa, A.; Kurosawa, Y.; Fujita  
Nucleic Acids Res. 15, 6733, 1987  
A;Title: Isolation of a full-length cDNA clone encoding human tyrosine hydroxylase type  
A;Reference number: I38340; MUID:87316931; PMID:2888085  
A;Accession: I38340  
A;Status: preliminary; translated from GB/EMBL/DBJ  
A;Molecule type: mRNA  
A;Residues: 1-30,35-528 <KOB4>  
A;Cross-references: EMBL:Y00414; NID:G37126; PIDN:CAA68472.1; PID:G37127  
R;Gins, E.I.; Rehavi, M.; Martin, B.M.; Weller, M.; O'Malley, K.L.; LaMarca, M.E.; McAL  
J. Biol. Chem. 263, 7406-7410, 1988  
A;Title: Expression of human tyrosine hydroxylase cDNA in invertebrate cells using a bac  
A;Reference number: I55282; MUID:88213428; PMID:2896667  
A;Accession: I55282  
A;Status: preliminary; translated from GB/EMBL/DBJ  
A;Molecule type: mRNA  
A;Residues: 1-30,62-64 <GIN1>  
A;Cross-references: GB:M20911; NID:G339636; PIDN:AAA61167.1; PID:G339637  
A;Accession: I70056  
A;Status: preliminary; translated from GB/EMBL/DBJ  
A;Molecule type: mRNA  
A;Residues: 1-34,62-64 <GIN2>  
A;Cross-references: GB:M20912; NID:G339642; PIDN:AAA61168.1; PID:G339643  
C;Comment: The expression of the four distinct proteins produced by alternate splicing v  
C;Genetics:  
A;Gene: GDB:TH  
A;Cross-references: GDB:119612; OMIM:191290  
A;Map position: 11p15.5-11p15.5  
A;Introns: 34/3; 61/3; 135/3  
A;Note: the list of introns is incomplete  
C;Function:  
A;Description: catalyzes the 3'-hydroxylation of tyrosine to 3',4'-dihydroxyphenylalanin  
A;Pathway: catecholamine biosynthesis  
A;Note: this is the rate-limiting step in catecholamine biosynthesis  
C;Superfamily: phenylalanine 4-monooxygenase  
C;Keywords: alternative splicing; bioppterin; catecholamine biosynthesis; iron; metallopro  
F;1-528/Product: tyrosine 3-monooxygenase, splice form 4 #status predicted <NA4>  
F;1-34,62-528/Product: tyrosine 3-monooxygenase, splice form 2 #status predicted <NA2>  
F;1-30,35-528/Product: tyrosine 3-monooxygenase, splice form 3 #status predicted <NA3>  
F;1-30,62-528/Product: tyrosine 3-monooxygenase, splice form 1 #status predicted <NA1>  
F;8/Binding site: phosphate (Thr) (covalent) (by unidentified kinase) #status predicted  
F;19/Binding site: phosphate (Ser) (covalent) (by calmodulin-dependent kinase) #status p  
F;71,183/Binding site: phosphate (Ser) (covalent) (by cAMP-dependent kinase) #status pre  
F;361,366,406/Binding site: iron (His, His, Glu) #status predicted  
  
Query Match 2.9%; Score 7; DB 1; Length 528;  
Best Local Similarity 100.0%; Pred. No. 90;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
Qy 121 EAAAVQP 127  
Db 446 EAAAVQP 452  
  
RESULT 50  
D43719  
N;Alternate names: 62K chain - Proteus mirabilis  
C;Species: Proteus mirabilis  
C;Date: 03-Mar-1993 #sequence\_revision 02-Dec-1994 #text\_change 05-Jan-2003  
C;Accession: D43719  
R;Jones, B.D.; Mobley, H.L.T.  
J. Bacteriol. 171, 6414-6422, 1989  
A;Title: Proteus mirabilis urease: nucleotide sequence determination and comparison with  
A;Reference number: A43719; MUID:90078080; PMID:2687233  
A;Accession: D43719  
A;Molecule type: DNA  
A;Residues: 1-567 <JON>  
A;Cross-references: GB:M31834; NID:G150914; PIDN:AAA25669.1; PID:G150918

C;Genetics:  
A;Gene: ureC  
A;Complex: trimer of heterotrimeric subunits consisting of 11K (gamma), 12K (beta), and  
C;Function:  
A;Description: catalyzes the hydrolysis of one molecule of urea to carbon dioxide and t  
C;Superfamily: urease, alpha subunit; urease 62K chain homology  
C;Keywords: heterotrimer; hydrolase; metalloprotein; nickel  
F;4-550/Domain: urease 62K chain homology <U62>  
F;134,136,217,360/Binding site: nickel 2 (His, His, Lys, Asp) #status predicted  
F;217,246,272/Binding site: nickel 1 (Lys, His, His) #status predicted  
F;217/Binding site: carbon dioxide (Lys) (covalent) (by urease activase) #status predic  
F;219,330/Active site: His #status predicted  
  
Query Match 2.9%; Score 7; DB 1; Length 567;  
Best Local Similarity 100.0%; Pred. No. 95;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
Qy 66 EAGVPEK 72  
Db 498 EAGVPEK 504  
  
RESULT 51  
S74477  
C;Superfamily: hypothetical protein slr1114 - Synechocystis sp. (strain PCC 6803)  
A;Species: Synechocystis sp.  
A;Variety: PCC 6803  
C;Date: 25-Apr-1997 #sequence\_revision 25-Apr-1997 #text\_change 20-Jun-2000  
C;Accession: S74477  
R;Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.; Miyajima, N.  
K. Okumura, S.; Shimpo, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yasud  
DNA Res. 3, 109-136, 1996  
A;Title: Sequence analysis of the genome of the unicellular cyanobacterium Synechocysti  
S.  
A;Reference number: S74322; MUID:97061201; PMID:8905231  
A;Accession: S74477  
A;Status: nucleic acid sequence not shown; translation not shown  
A;Molecule type: DNA  
A;Residues: 1-585 <KAN>  
A;Cross-references: EMBL:D90899; GB:AB001339; NID:G1651650; PIDN:BAAL6629.1; PID:G16517  
A;Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1996  
C;Superfamily: Synechocystis hypothetical protein slr1114  
  
Query Match 2.9%; Score 7; DB 2; Length 585;  
Best Local Similarity 100.0%; Pred. No. 98;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
Qy 161 IAIAGGI 167  
Db 512 IAIAGGI 518  
  
RESULT 52  
S77707  
phosphoribosylaminoimidazolecarboxamide formyltransferase (EC 2.1.2.3) - yeast (Sacchar  
N;Alternate names: 5-aminimidazole-4-carboxamide ribotide transformylase; protein L175  
C;Species: Saccharomyces cerevisiae  
C;Date: 02-May-1997 #sequence\_revision 02-May-1997 #text\_change 18-Jun-1999  
C;Accession: S77707; S64855  
R;Tibbets, A.S.; Appling, D.R.  
submitted to the EMBL Data Library, June 1996  
A;Description: Isolation and characterization of two yeast genes encoding 5-aminimidaz  
A;Reference number: S77707  
A;Accession: S77707  
A;Molecule type: DNA  
A;Residues: 1-591 <TIB>  
A;Cross-references: EMBL:U62402; NID:G1480727; PIDN:AA57774.1; PID:G1480728  
R;Oberman, B.; Piravandi, E.; Rinke, M.  
submitted to the Protein Sequence Database, May 1996  
A;Reference number: S64845  
A;Accession: S64855  
A;Molecule type: DNA  
A;Residues: 1-545, 'TFIELC', 553, 'L', 555 <OBE>



C>Date: 13-Sep-1996 #sequence\_revision 13-Sep-1996 #text\_change 18-Jun-1999  
C/Accession: I51283  
R:Nakamura, H.; Iashiro, K.; Nakamura, T.; Shiokawa, K.  
Mech. Dev. 49, 123-131, 1995  
A>Title: Molecular cloning of Xenopus HGF cDNA and its expression studies in Xenopus ear  
A/Reference number: I51283; MUID:95267690; PMID:7748783  
A/Accession: I51283  
A/Status: preliminary; translated from GB/EMBL/DBJ  
A/Molecule type: mRNA  
A/Residues: 1-710 <NAK>  
A/Cross-references: GB:I577422; NID:9989932; PIDN:AA834354.1; PID:9989933  
A/Note: the authors' translation for residue 458 (Thr) is inconsistent with the nucleotide sequence  
C/Complex: disulfide-bonded heterodimer of chains derived from the same precursor  
C/Function:  
A/Description: stimulates mitosis of hepatocytes and other cells  
A/Note: Does not have proteinase activity  
C/Subfamily: hepatocyte growth factor; kringle homology; trypsin homology  
C/Keywords: duplication; glycoprotein; growth factor; heterodimer; kringle  
F:42-477/Domain: hepatocyte growth factor #status predicted <NAK>  
F:115-193/Domain: hepatocyte growth factor alpha chain #status predicted <ACH>  
F:198-275/Domain: kringle homology <KR1>  
F:289-367/Domain: kringle homology <KR2>  
F:375-453/Domain: kringle homology <KR4>  
F:478-709/Domain: hepatocyte growth factor beta chain #status predicted <BCH>  
F:478-709/Domain: hepatocyte growth factor <TRY>  
F:52,128,281,322,379,550,637,666/Binding site: carbohydrate (Asn) (covalent) #status predicted  
F:470-588/Disulfide bonds: #status predicted  
Query Match 2.9%; Score 7; DB 1; Length 710;  
Best Local Similarity 100.0%; Pred. No. 1.2e+02; Indels 0; Gaps 0;  
Matches 7; Conservative 0; Mismatches 0;  
QY 175 RGRDLKE 181  
DB 155 RGRDLKE 161  
RESULT 58  
T38359  
cullin 3 homolog - fission yeast (Schizosaccharomyces pombe)  
C/Species: Schizosaccharomyces pombe  
C/Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 03-Dec-1999  
C/Accession: T38359  
R/Skelton, J.; Churcher, C.M.; Barrell, B.G.; Rajandream, M.A.; Walsh, S.V.; Wood, V.  
submitted to the EMBL Data Library, September 1995  
A/Reference number: 221788  
A/Status: preliminary; translated from GB/EMBL/DBJ  
A/Molecule type: DNA  
A/Residues: 1-785 <SKE>  
A/Cross-references: EMBL:Z54142; PIDN:CRA90847.1; GSPDB:GN00066; SPDB:SPAC24H6.03  
A/Experimental source: strain 972h-; cosmid c24H6  
C/Genetics:  
A/Gene: pcu3; SPDB:SPAC24H6.03  
A/Map position: 1  
A/Introns: 14/3; 476/2; 513/1; 534/3; 592/3; 729/2  
Query Match 2.9%; Score 7; DB 2; Length 785;  
Best Local Similarity 100.0%; Pred. No. 1.3e+02; Indels 0; Gaps 0;  
Matches 7; Conservative 0; Mismatches 0;  
QY 82 ETTSQL 88  
DB 353 ETTSQL 359  
RESULT 59  
S62405  
hypothetical protein SPAC24H6.03 - fission yeast (Schizosaccharomyces pombe)  
C/Species: Schizosaccharomyces pombe  
C/Date: 16-May-1996 #sequence\_revision 13-Mar-1997 #text\_change 31-Oct-1997  
C/Accession: S62405

R/Skelton, J.; Churcher, C.M.  
submitted to the EMBL Data Library, September 1995  
A/Reference number: S62402  
A/Accession: S62405  
A/Status: preliminary  
A/Molecule type: DNA  
A/Residues: 1-798 <SKE>  
A/Cross-references: EMBL:Z54142; NID:9984697; PID:9984700  
C/Genetics:  
A/Map position: 1L  
A/Introns: 14/3; 526/1; 547/3; 605/3; 742/2  
Query Match 2.9%; Score 7; DB 2; Length 798;  
Best Local Similarity 100.0%; Pred. No. 1.3e+02; Indels 0; Gaps 0;  
Matches 7; Conservative 0; Mismatches 0;  
QY 82 ETTSQL 88  
DB 353 ETTSQL 359  
RESULT 60  
A45082  
neutrotrophic receptor ror1 precursor - human  
N/Contains: protein-tyrosine kinase (EC 2.7.1.112)  
C/Species: Homo sapiens (man)  
C/Date: 30-Sep-1993 #sequence\_revision 30-Sep-1993 #text\_change 19-May-2000  
C/Accession: A45082  
R/Masiakowski, P.; Carroll, R.D.  
J. Biol. Chem. 267, 26181-26190, 1992  
A/Title: A novel family of cell surface receptors with tyrosine kinase-like domain.  
A/Reference number: A45082; MUID:93100347; PMID:1334494  
A/Accession: A45082  
A/Molecule type: mRNA  
A/Residues: 1-937 <MAS>  
A/Cross-references: GB:M97675; NID:G337464; PIDN:AAA60275.1; PID:G337465  
A/Note: sequence extracted from NCBI backbone (NCBIP:120916)  
C/Genetics:  
A/Gene: GDB:NTRK1  
A/Cross-references: GDB:I36453  
A/Map position: 6p21-6p21  
C/Supfamily: neurotrophic receptor ror; immunoglobulin homology; kringle homology; pr  
C/Keywords: ATP; glycoprotein; kringle; phosphotransferase; transmembrane protein; tyro  
F:1-23/Domain: signal sequence #status predicted <Sig>  
F:24-937/Product: neurotrophic receptor ror1 #status predicted <MAT>  
F:72-133/Domain: immunoglobulin homology <IMW>  
F:313-391/Domain: kringle homology <KEG>  
F:404-425/Domain: transmembrane #status predicted <TML>  
F:471-753/Domain: protein kinase homology <KIN>  
F:47,66,184,315/Binding site: carbohydrate (Asn) (covalent) #status predicted  
Query Match 2.9%; Score 7; DB 2; Length 937;  
Best Local Similarity 100.0%; Pred. No. 1.5e+02; Indels 0; Gaps 0;  
Matches 7; Conservative 0; Mismatches 0;  
QY 46 HSYCRNP 52  
DB 359 HSYCRNP 365  
RESULT 61  
G70951  
probable ATP-dependent DNA helicase - Mycobacterium tuberculosis (strain H37RV)  
C/Species: Mycobacterium tuberculosis  
C/Date: 17-Jul-1998 #sequence\_revision 17-Jul-1998 #text\_change 22-Oct-1999  
C/Accession: G70951  
R/Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.  
; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.  
Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.  
Nature 393, 537-544, 1998  
A/Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.  
A/Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome  
A/Reference number: A70500; MUID:98295987; PMID:9634230

A:Accession: G70951  
A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-1101 <COL>  
A:Cross-references: GB:AL021646; GB:AL123456; NID:g3242278; PIDN:CAA16666.1; PID:e124879  
A:Experimental source: strain H37RV  
C:Genetics:  
A:Gene: RV3201C

Query Match 2.9%; Score 7; DB 2; Length 1101;  
Best Local Similarity 100.0%; Pred. No. 1.7e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 36 ASAPVSG 42  
|||||  
DB 594 ASAPVSG 600

RESULT 62  
S4247  
parasporal crystal protein cry9Cal [validated] - Bacillus thuringiensis  
N:Alternate names: parasporal crystal protein cryIH  
C:Species: Bacillus thuringiensis  
C:Date: 01-Dec-2000 #sequence\_revision 01-Dec-2000 #text\_change 01-Dec-2000  
C:Accession: A59350; S49247  
R:Lambert, B.; Buysse, L.; Decock, C.; Janssens, S.; Piens, C.; Saey, B.; Seurinck, J.; V  
Appl. Environ. Microbiol. 62, 80-86, 1996  
A:Title: A Bacillus thuringiensis insecticidal crystal protein with a high activity agai  
A:Reference number: A59350; MUID:96141404; PMID:8572715  
A:Accession: A59350  
A:Molecule type: DNA  
A:Residues: 1-1157 <LAM>  
A:Cross-references: EMBL:Z37527; NID:G547554; PIDN:CAA85764.1; PID:G547556  
A:Experimental source: serovar tolworthi  
C:Comment: This parasporal crystal protein, active against corn borer and other insects,  
C:Superfamily: parasporal crystal protein  
C:Keywords: delta-endotoxin

Query Match 2.9%; Score 7; DB 1; Length 1157;  
Best Local Similarity 100.0%; Pred. No. 1.8e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 36 ASAPVSG 42  
|||||  
DB 523 ASAPVSG 529

RESULT 63  
T16859  
hypothetical protein T13C2.5 - Caenorhabditis elegans  
C:Species: Caenorhabditis elegans  
C:Date: 20-Sep-1999 #sequence\_revision 20-Sep-1999 #text\_change 17-Nov-2000  
C:Accession: T16859  
R:Du, Z.  
submitted to the EMBL Data Library, November 1995  
A:Description: The sequence of C. elegans cosmid T13C2.  
A:Reference number: Z18591  
A:Accession: T16859  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-1296 <DUZ>  
A:Cross-references: EMBL:U40030; NID:g1055164; PID:g1055165; PIDN:AAA81133.1; CESP:T13C2  
C:Genetics:  
A:Gene: CESP:T13C2.5  
A:Introns: 62/3; 93/3; 124/3; 142/1; 212/1; 293/1; 336/2; 406/3; 437/1; 508/1; 584/1; 62

Query Match 2.9%; Score 7; DB 2; Length 1296;  
Best Local Similarity 100.0%; Pred. No. 2e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 74 PCEDLRC 80  
|||||  
DB 216 PCEDLRC 222

## RESULT 64

H70846

hypothetical glycine-rich protein RV3345c - Mycobacterium tuberculosis (strain H37RV)  
C:Species: Mycobacterium tuberculosis  
C:Date: 17-Jul-1998 #sequence\_revision 17-Jul-1998 #text\_change 20-Jun-2000  
C:Accession: H70846  
R:Coile, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, J.  
; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.  
Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.  
Nature 393, 537-544, 1998  
A:Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.  
A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome  
A:Reference number: A70500; MUID:98295987; PMID:9634230  
A:Accession: H70846  
A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-1538 <COL>  
A:Cross-references: GB:AL021841; GB:AL123456; NID:g3261517; PIDN:CAA17117.1; PID:g32615  
A:Experimental source: strain H37RV  
C:Genetics:  
A:Gene: RV3345c  
C:Superfamily: collagen alpha 1 (IV) chain

Query Match 2.9%; Score 7; DB 2; Length 1538;  
Best Local Similarity 100.0%; Pred. No. 2.3e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 235 GQAGTPG 241

|||||

DB 776 GQAGTPG 782

|||||

## RESULT 65

T30313

chemotaxis protein homolog - Pseudomonas aeruginosa  
C:Species: Pseudomonas aeruginosa  
C:Date: 22-Oct-1999 #sequence\_revision 22-Oct-1999 #text\_change 22-Oct-1999  
C:Accession: T30313  
R:Whitchurch, C.B.; Young, M.D.; Hobbs, M.; Mattick, J.S.  
submitted to the EMBL Data Library, November 1996  
A:Description: Pseudomonas aeruginosa chemotactic transduction genes pill, chpA chpB an  
A:Reference number: Z20819  
A:Accession: T30313  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-1638 <WHI>  
A:Cross-references: EMBL:U79580; NID:g3241967; PID:g3241969; PIDN:AAC23931.1  
C:Genetics:  
A:Note: chpA

Query Match 2.9%; Score 7; DB 2; Length 1638;  
Best Local Similarity 100.0%; Pred. No. 2.4e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 29 LDAQSGL 35

|||||

DB 919 LDAQSGL 925

|||||

## RESULT 66

E83594

still frameshift probable component of chemotactic signal transduction system PA0413 [1]  
C:Species: Pseudomonas aeruginosa  
C:Date: 15-Sep-2000 #sequence\_revision 15-Sep-2000 #text\_change 31-Dec-2000  
C:Accession: E83594  
R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warren, P.; Hickey, M.J.; B.  
adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; LatBis, K.; Lim  
; Lory, S.; Olson, M.V.  
Nature 406, 959-964, 2000  
A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic path  
A:Reference number: A82950; MUID:2043737; PMID:10984043

A;Accession: E83594  
 A;Status: preliminary  
 A;Molecule type: DNA  
 A;Residues: 1-2472 <STO>  
 A;Cross-references: GB:AE004478; GB:AE004091; NID:g9946261; PIDN:AA03802.1; GSPDB:GN001  
 A;Experimental source: strain PA01  
 C;Genetics:  
 A;Gene: PA0413

Query Match 2.9%; Score 7; DB 2; Length 2472;  
 Best Local Similarity 100.0%; Pred. No. 3.4e+02; Mismatches 0; Indels 0; Gaps 0;  
 Matches 7; Conservative 0;

QY 29 LDAQSL 35  
 |||||  
 Db 1753 LDAQSL 1759

## RESULT 67

GNVVKV  
 genome polypeptide - Kunjin virus (strain MRM61C)  
 N;Contains: capsid protein C; envelope protein E; membrane protein M; nonstructural protein NS4a; nonstructural protein NS4b; nonstructural protein NS5  
 C;Species: Kunjin virus  
 C;Date: 30-Sep-1989 #sequence\_revision 30-Sep-1989 #text\_change 19-Jan-2001  
 C;Accession: A28697  
 R;Coila, G.; Parker, M.D.; Speight, G.; Byrne, M.E.; Westaway, E.G.  
 J. Gen. Virol. 69, 1-21, 1988  
 A;Title: Nucleotide and complete amino acid sequences of Kunjin virus: definitive gene C  
 A;Reference number: A28697; MUID:88089524; PMID:2826659  
 A;Accession: A28697  
 A;Molecule type: genomic RNA  
 A;Residues: 1-3433 <COI>

A;Cross-references: GB:D00246; NID:g221966; PIDN:BAA00176.1; PID:g221967  
 C;Superfamily: yellow fever virus genome polypeptide  
 C;Keywords: ATP; capsid protein; envelope protein; membrane protein; nonstructural protein  
 F;2-123/Product: capsid protein C #status predicted <CPC>  
 F;124-290/Product: membrane protein M precursor #status predicted <MPP>  
 F;124-215/Domain: nonterminal signal sequence #status predicted <SIG>  
 F;126-290/Product: membrane protein M #status predicted <MEM>  
 F;291-791/Product: envelope protein E #status predicted <EPE>  
 F;792-1143/Product: nonstructural protein NS1 #status predicted <NS1>  
 F;1144-1374/Product: nonstructural protein NS2a #status predicted <NS2a>  
 F;1375-1505/Product: nonstructural protein NS2b #status predicted <NS2b>  
 F;1506-2124/Product: nonstructural protein NS3 #status predicted <NS3>  
 F;1699-1706/Region: nucleotide-binding motif A (P-loop)  
 F;1786-1791/Region: nucleotide-binding motif B  
 F;1790-1793/Region: DEAH motif  
 F;2125-2273/Product: nonstructural protein NS4a #status predicted <NS4a>  
 F;2274-2528/Product: nonstructural protein NS4b #status predicted <NS4b>  
 F;2529-3433/Product: nonstructural protein NS5 #status predicted <NS5>

Query Match 2.9%; Score 7; DB 1; Length 3433;  
 Best Local Similarity 100.0%; Pred. No. 4.6e+02; Mismatches 0; Indels 0; Gaps 0;  
 Matches 7; Conservative 0;

QY 21 PGLRCLN 27  
 |||||  
 Db 1307 PGLRCLN 1313

## RESULT 68

E86077  
 hypothetical protein Z5430 [imported] - Escherichia coli (strain O157:H7, substrain EDL9  
 C;Species: Escherichia coli  
 C;Date: 16-Feb-2001 #sequence\_revision 16-Feb-2001 #text\_change 14-Sep-2001  
 C;Accession: E86077  
 R;Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew  
 Miller, L.; Grotbeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamousis, K.; Apodaca,  
 Nature 409, 529-533, 2001  
 A;Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.  
 A;Reference number: A85480; MUID:21074935; PMID:11206551  
 A;Accession: E86077

A;Status: preliminary  
 A;Molecule type: DNA  
 A;Residues: 1-38 <STO>  
 A;Cross-references: GB:AE005174; NID:g12518780; PIDN:AA059081.1; GSPDB:GN00145; UWGP:Z5  
 A;Experimental source: strain O157:H7, substrain EDL933  
 C;Genetics:  
 A;Gene: Z5430

Query Match 2.5%; Score 6; DB 2; Length 38;  
 Best Local Similarity 100.0%; Pred. No. 95; Mismatches 0; Indels 0; Gaps 0;  
 Matches 6; Conservative 0;

QY 68 GVPEKR 73  
 |||||  
 Db 25 GVPEKR 30

## RESULT 69

A42064  
 lactam utilization protein lamA - Emericella nidulans (fragment)  
 C;Species: Emericella nidulans, Aspergillus nidulans  
 C;Date: 31-Dec-1993 #sequence\_revision 31-Dec-1993 #text\_change 17-Mar-1999  
 C;Accession: A42064  
 R;Richardson, I.B.; Katz, M.E.; Hynes, M.J.  
 Mol. Cell. Biol. 12, 337-346, 1992

A;Title: Molecular characterization of the lam locus and sequences involved in regulati  
 A;Reference number: A42064; MUID:92107186; PMID:1729609  
 A;Accession: A42064

A;Status: preliminary  
 A;Molecule type: DNA  
 A;Residues: 1-41 <RIC>  
 A;Cross-references: GB:M77283

Query Match 2.5%; Score 6; DB 2; Length 41;  
 Best Local Similarity 100.0%; Pred. No. 1e+02; Mismatches 0; Indels 0; Gaps 0;  
 Matches 6; Conservative 0;

QY 17 TSPAPG 22  
 |||||  
 Db 33 TSPAPG 38

## RESULT 70

B64323  
 hypothetical protein MJ0185 - Methanococcus jannaschii  
 C;Species: Methanococcus jannaschii  
 C;Date: 13-Sep-1996 #sequence\_revision 13-Sep-1996 #text\_change 22-Oct-1999  
 C;Accession: B64323  
 R;Bult, C.J.; White, O.; Olsen, G.J.; Zhou, L.; Fleischmann, R.D.; Sutton, G.G.; Blake,  
 Reich, C.I.; Overbeek, R.; Kirkness, E.F.; Weinstock, K.G.; Merrick, J.M.; Glodek, A.  
 rson, J.D.; Sadow, P.W.; Hanna, M.C.; Cotton, M.D.; Roberts, K.M.; Hurst, M.A.  
 Science 273, 1058-1073, 1996

A;Authors: Kaine, B.P.; Borodovsky, M.; Klenk, H.P.; Fraser, C.M.; Smith, H.O.; Woese,  
 A;Title: Complete genome sequence of the methanogenic archaeon, Methanococcus jannaschii  
 A;Reference number: A64300; MUID:96337999; PMID:8688087  
 A;Accession: B64323

A;Status: preliminary; nucleic acid sequence not shown; translation not shown  
 A;Molecule type: DNA

A;Residues: 1-49 <BUL>  
 A;Cross-references: GB:U67475; GB:L77117; NID:gl590930; PIDN:AA059081.1; PID:gl498959;  
 C;Genetics:  
 A;Map position: FOR183886-184035

Query Match 2.5%; Score 6; DB 2; Length 49;  
 Best Local Similarity 100.0%; Pred. No. 1.2e+02; Mismatches 0; Indels 0; Gaps 0;  
 Matches 6; Conservative 0;

QY 159 IIIAIG 164  
 |||||  
 Db 30 IIIAIG 35

## RESULT 71

H83236  
cold acclimation protein B PA3266 [imported] - *Pseudomonas aeruginosa* (strain PA01)  
C:Species: *Pseudomonas aeruginosa*  
C>Date: 15-Sep-2000 #sequence\_revision 15-Sep-2000 #text\_change 15-Sep-2003  
C:Accession: H83236  
R:Stover, C.K.; Pham, X.O.; Erwin, A.L.; Mizoguchi, S.D.; Warren, P.; Hickey, M.J.; Berman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; Lim, J.; Lory, S.; Olson, M.V.  
Nature 406, 959-964, 2000  
A:Title: Complete genome sequence of *Pseudomonas aeruginosa* PA01, an opportunistic pathogen  
A:Reference number: A82950; MUID:20437337; PMID:10984043  
A:Accession: H83236  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-69 <STO>  
A:Cross-references: GB:AE004749; GB:AE004091; NID:g9949398; PIDN:AAG0654.1; GSPDB:GN00138  
A:Experimental source: strain PA01  
C:Genetics:  
A:Gene: capB; PA3266  
C:Superfamily: cold shock protein; cold shock domain homology

Query Match 2.5%; Score 6; DB 2; Length 69;  
Best Local Similarity 100.0%; Pred. No. 1.6e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 104 ADEVQV 109  
|||||  
Db 63 ADEVQV 68

RESULT 72  
B84284  
hypothetical protein Vng1289h [imported] - *Halobacterium* sp. NRC-1  
C:Species: *Halobacterium* sp. NRC-1  
C>Date: 02-Feb-2001 #sequence\_revision 02-Feb-2001 #text\_change 02-Feb-2001  
C:Accession: B84284  
R:Ng, W.V.; Kennedy, S.P.; Mahairas, G.G.; Berquist, B.; Pan, M.; Shukla, H.D.; Lasky, S.; Leithauser, B.; Keller, K.; Cruz, R.; Danson, M.J.; Hough, D.W.; Maddocks, D.G.; Jablo, J.; Jung, K.H.; Alam, M.; Freitas, T.  
Proc. Natl. Acad. Sci. U.S.A. 97, 12176-12181, 2000  
A:Authors: Hou, S.; Daniels, C.J.; Dennis, P.P.; Omer, A.D.; Ebhardt, H.; Lowe, T.M.; Li, J.  
A:Title: Genome sequence of *Halobacterium* species NRC-1.  
A:Reference number: A84160; MUID:20504483; PMID:11016950  
A:Accession: B84284  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-71 <STO>  
A:Cross-references: GB:AE004437; NID:g10580810; PIDN:AAG19638.1; GSPDB:GN00138  
C:Genetics:  
A:Gene: Vng1289H

Query Match 2.5%; Score 6; DB 2; Length 71;  
Best Local Similarity 100.0%; Pred. No. 1.6e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 66 EAGVPE 71  
|||||  
Db 47 EAGVPE 52

RESULT 73  
AD1043  
transcription regulatory protein cII [imported] - *Salmonella enterica* subsp. *enterica* serovar Typhi  
C:Species: *Salmonella enterica* subsp. *enterica* serovar Typhi  
A:Note: this species has also been called *Salmonella typhi*  
C>Date: 09-Nov-2001 #sequence\_revision 09-Nov-2001 #text\_change 18-Nov-2002  
C:Accession: AD1043  
R:Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Churcher, C.; Mouton, S.; O'Garra, P.; S.; Mouton, S.; O'Garra, P.  
Nature 413, 848-852, 2001  
A:Authors: Parry, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.; A:Title: Complete genome sequence of a multiple drug resistant *Salmonella enterica* serovar

A:Reference number: AB0502; MUID:21534947; PMID:11677608  
A:Accession: AD1043  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-73 <PAR>  
A:Cross-references: GB:AL513382; PIDN:CAD06793.1; PID:g16505443; GSPDB:GN00176  
C:Genetics:  
A:Gene: cII  
C:Superfamily: phage D108 DNA-binding protein

Query Match 2.5%; Score 6; DB 2; Length 73;  
Best Local Similarity 100.0%; Pred. No. 1.7e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 32 QSGLAS 37  
|||||  
Db 28 QSGLAS 33

RESULT 74  
TI4887  
hypothetical protein yjCB - *Salmonella typhimurium*  
C:Species: *Salmonella typhimurium*  
C>Date: 20-Sep-1999 #sequence\_revision 20-Sep-1999 #text\_change 20-Sep-1999  
C:Accession: TI4887  
R:Wong, K.K.; McClelland, M.; Stillwell, L.C.; Sisk, E.C.; Thurston, S.J.; Saffer, J.D.  
Infect. Immun. 66, 3365-3371, 1998  
A:Title: Identification and sequence analysis of a 27-kilobase chromosomal fragment containing typhimurium LT2.  
A:Reference number: Z18249; MUID:98298059; PMID:9632606  
A:Accession: TI4887  
A:Status: preliminary; translated from GB/EMBL/DDJ  
A:Molecule type: DNA  
A:Residues: 1-73 <WON>  
A:Cross-references: EMBL:AF060869; NID:g3323584; PID:g3323604; PIDN:AAC26656.1  
C:Genetics:  
A:Gene: yjCB  
A:Map position: 92 min

Query Match 2.5%; Score 6; DB 2; Length 73;  
Best Local Similarity 100.0%; Pred. No. 1.7e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 152 LGITMM 157  
|||||  
Db 39 LGITMM 44

RESULT 75  
D82753  
hypothetical protein XF0871 [imported] - *Xylella fastidiosa* (strain 9a5c)  
C:Species: *Xylella fastidiosa*  
C>Date: 18-Aug-2000 #sequence\_revision 20-Aug-2000 #text\_change 20-Aug-2000  
C:Accession: D82753  
R:Anonymous, The Xylella fastidiosa Consortium of the Organization for Nucleotide Sequencing  
Nature 406, 151-157, 2000  
A:Title: The genome sequence of the plant pathogen *Xylella fastidiosa*.  
A:Reference number: A82515; MUID:20365717; PMID:10910347  
A:Note: for a complete list of authors see reference number A59328 below  
A:Accession: D82753  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-74 <SIM>  
A:Cross-references: GB:AE003926; GB:AE003849; NID:g9105771; PIDN:AAF83681.1; GSPDB:GN00138  
A:Experimental source: strain 9a5c  
R:Simpson, A.J.G.; Reinach, F.C.; Arruda, P.; Abreu, F.A.; Acencio, M.; Alvaranga, R.; Briones, M.R.S.; Bueno, M.R.P.; Camargo, A.A.; Camargo, L.E.A.; Carraro, D.M.; Carrer, as-Neto, E.; Docena, C.; El-Dorri, H.; Facincani, A.P.; Ferreira, A.J.S.  
submitted to GenBank, June 2000  
A:Authors: Ferreira, V.C.A.; Ferro, J.A.; Fraga, J.S.; Franco, S.C.; Franco, M.C.; Froh, J.D.; Junqueira, M.L.; Kemper, E.L.; Kitajima, J.P.; Krieger, J.E.; Kuramae, E.E.; Laig, chado, M.A.; Madeira, A.M.B.N.; Madeira, H.M.F.; Marino, C.L.; Marques, M.V.; Martins, A:Authors: Martins, E.M.F.; Matsukuma, A.Y.; Menck, C.F.M.; Miracca, E.C.; Miyaki, C.Y.

F.G.; Nunes, L.R.; Oliveira, M.A.; de Oliveira, M.C.; de Oliveira, R.C.; Palmieri, D.A.  
Rodrigues, V.; Rosa, A.J. de M.; de Rosa Jr., V.E.; de Sa, R.G.; Santelli, R.V.; Sawasak  
A;Authors: da Silva, A.C.R.; da Silva, F.R.; da Silva, A.M.; Silva Jr., W.A.; da Silveir  
M.; Tsubako, M.H.; Vallada, H.; Van Sluys, M.A.; Verjovski-Almeida, S.; Vettore, A.L.; Z  
A;Reference number: A59328  
A;Contents: annotation  
C;Genetics:  
A;Gene: XF0871

Query Match 2.5%; Score 6; DB 2; Length 74;  
Best Local Similarity 100.0%; Pred. No. 1.7e-02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 134 RVRMNS 139  
| | | | |  
Db 69 RVRMNS 74

Search completed: March 17, 2004, 07:26:50  
Job time : 25 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 17, 2004, 07:09:14 ; Search time 18 seconds  
(without alignments)  
700.054 Million cell updates/sec

Title: US-10-057-951-2\_COPY\_22\_263  
Perfect score: 242  
Sequence: 1 SGCCFWNDHLYREDQTSFA.....PVDQEGSTPLMGQAGTGA 242

Scoring table: OLIGO  
Gapop 60.0 , Gapext 60.0

Searched: 141681 seqs, 52070155 residues

Word size : 0

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Listing first 100 summaries

Database : SwissProt\_42.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	8	3.3	429	PURA_THIFE	P52151 thibobacillu
2	8	3.3	716	HGFL_MOUSE	P26928 mus musculu
3	8	3.3	771	PCRA_MYCTU	P71561 mycobacteri
4	7	2.9	96	CH10_AMOPS	P26001 amoeba prot
5	7	2.9	154	YL66_ARCFU	Q28116 archaeoglob
6	7	2.9	281	STO4_CABEL	Q22165 caenorhabd
7	7	2.9	301	SC14_CANAL	P46250 candida alb
8	7	2.9	335	GPDA_HAEIN	P43798 haemophilus
9	7	2.9	335	Y780_METJA	Q58190 methanococc
10	7	2.9	336	GPDA_HAEIDU	Q7vna0 haemophilus
11	7	2.9	388	XYLR_LACPE	P21940 lactobacill
12	7	2.9	398	AAT_FSEAE	P72173 pseudomonas
13	7	2.9	402	PURA_BRUAB	P52004 brucella ab
14	7	2.9	429	PURA_BRUME	Q8yiu1 brucella me
15	7	2.9	432	PURA_RHILO	Q98f97 rhizobium l
16	7	2.9	464	SYER_CORBU	Q83816 coxiella bu
17	7	2.9	479	FTSA_FORGI	Q07827 porphyromon
18	7	2.9	512	EMRB_ECOLI	P27304 escherichia
19	7	2.9	528	TY3H_HUNAN	P07101 homo sapien
20	7	2.9	567	URE1_PROVI	P17086 proteus mir
21	7	2.9	591	PU91_YEAST	P54113 s bifunctio
22	7	2.9	592	PU92_YEAST	P38009 s bifunctio
23	7	2.9	619	LT65_ARATH	Q04980 arabidopsis
24	7	2.9	661	UAS3_HUNAN	P57075 homo sapien
25	7	2.9	769	LEM3_SHEEP	P98109 ovis aries
26	7	2.9	785	CUL3_SCHPO	Q01973 homo sapien
27	7	2.9	937	ROR1_HUMAN	Q92139 mus musculu
28	7	2.9	937	ROR1_MOUSE	P54362 drosophila
29	7	2.9	1034	ADD_DROME	P45733 bacillus th
30	7	2.9	1157	CSCA_BACTO	P14335 k genome po
31	7	2.9	3433	POIG_KUNUM	P38095 emericella
32	6	2.5	41	LAMA_EMEI	Q57644 methanococc
33	6	2.5	49	Y185_METJA	

34	6	2.5	63	1	BD02_RAT	O88514 rattus norv
35	6	2.5	69	1	CSPA_PEEAE	P95459 pseudomonas
36	6	2.5	78	1	FY81_SYNEL	P50035 synecococc
37	6	2.5	81	1	ATPL_MYCLE	P45828 mycobacteri
38	6	2.5	81	1	ATPL_MYCTU	Q10598 mycobacteri
39	6	2.5	83	1	RS17_CHLMU	Q9pjm3 chlamydia m
40	6	2.5	83	1	RS17_CHLTR	P28545 chlamydia t
41	6	2.5	86	1	RS17_CHLPN	Q9z7r6 chlamydia p
42	6	2.5	93	1	IHF8_PASHA	P95519 pasteurrella
43	6	2.5	95	1	Y492_METJA	Q57915 methanococc
44	6	2.5	105	1	THIO_ALIAC	P80579 allicyclobac
45	6	2.5	106	1	RLA3_YEAST	P10622 saccharomyc
46	6	2.5	106	1	THCC_RHOER	P43493 rhodococcus
47	6	2.5	108	1	DBH_BORAD	Q46255 borrelia an
48	6	2.5	108	1	DBH_BORAF	Q57220 borrelia af
49	6	2.5	108	1	DBH_BORBU	Q57267 borrelia bu
50	6	2.5	108	1	DBH_BORJA	Q57153 borrelia ga
51	6	2.5	108	1	DBH_BORJA	Q45231 borrelia ja
52	6	2.5	108	1	DBH_BORTU	Q45722 borrelia tu
53	6	2.5	109	1	RLA1_SCHPO	P17476 schizosacch
54	6	2.5	109	1	RLA5_SCHPO	Q9uu78 schizosacch
55	6	2.5	109	1	RS17_HALN1	O24786 halobacteri
56	6	2.5	110	1	RLA3_SCHPO	P17477 schizosacch
57	6	2.5	111	1	RS17_HALMA	P12741 halosarcula
58	6	2.5	118	1	IAA4_SORBI	P81367 sorghum bic
59	6	2.5	121	1	YHAH_ECOLI	P42621 escherichia
60	6	2.5	122	1	RL7_LIBAS	P36247 liberibacte
61	6	2.5	124	1	NO5_VIGSA	Q41705 vicia sativ
62	6	2.5	130	1	CAL2_MOUSE	Q99mp3 mus musculu
63	6	2.5	133	1	RNPA_CORGL	O8n151 corynebacte
64	6	2.5	135	1	H32_TETPY	P15512 tetrahymena
65	6	2.5	135	1	H33_TETPH	P41553 tetrahymena
66	6	2.5	135	1	NO5_PEA	P25226 pisum sativ
67	6	2.5	136	1	RUVX_DEIRA	Q9rri2 deinococcus
68	6	2.5	139	1	RUVX_HABIN	P43981 haemophilus
69	6	2.5	140	1	RUVX_PASMU	Q9cix0 pasteurella
70	6	2.5	141	1	HBAA_TRICR	P10784 triturus cr
71	6	2.5	144	1	RIB1_PHOLE	Q01984 photobacter
72	6	2.5	148	1	MBB1_ECOLI	P08037 escherichia
73	6	2.5	148	1	Y222_TREPA	O83251 treponema p
74	6	2.5	150	1	YJGK_ECOLI	P39335 escherichia
75	6	2.5	153	1	SODC_NEUCR	P07509 neurospora
76	6	2.5	160	1	YQGC_BAGSU	P54486 bacillus su
77	6	2.5	169	1	PLMN_RAT	Q01177 rattus norv
78	6	2.5	173	1	LEPA_MYCHY	Q9zhs8 mycoplasma
79	6	2.5	175	1	PEIB_STRCH	P77949 streptomyce
80	6	2.5	176	1	Y225_MYCTU	Q50703 mycobacteri
81	6	2.5	178	1	Y229_MYCTU	O06246 mycobacteri
82	6	2.5	180	1	HBL0_PHYPA	Q9m630 physcomitre
83	6	2.5	185	1	ADML_RAT	P43145 rattus norv
84	6	2.5	187	1	RPE1_METJA	Q37840 methanococc
85	6	2.5	192	1	NIFZ_FRAHL	P46040 frankia aln
86	6	2.5	193	1	R122_ARATH	P36211 arabidopsis
87	6	2.5	201	1	Y802_NEIMA	Q9jvk3 neisseria m
88	6	2.5	204	1	PDGA_RAT	P28576 rattus norv
89	6	2.5	206	1	NEF_HVIEL	P04604 human immu
90	6	2.5	207	1	NEF_HVIND	P18801 human immu
91	6	2.5	211	1	PDGA_HUMAN	P04085 homo sapien
92	6	2.5	211	1	PDGA_MOUSE	P30033 mus musculu
93	6	2.5	212	1	ALKH_HAEIN	P44480 h putative
94	6	2.5	214	1	PSAE_YERPE	P31524 yersinia pe
95	6	2.5	215	1	GGL2_HUNAN	P36268 homo sapien
96	6	2.5	218	1	YXJL_BAGSU	P55184 bacillus su
97	6	2.5	219	1	CLN1_HUMAN	Q9bxu9 homo sapien
98	6	2.5	221	1	GTA1_CHICK	Q08392 gallus gall
99	6	2.5	222	1	GTA2_CHICK	Q08393 gallus gall
100	6	2.5	222	1	GTA3_RAT	P14942 rattus norv

ALIGNMENTS

RESULT 1



PURA THIFE  
ID PURA\_THIFE STANDARD; PRT; 429 AA.  
AC P52151;  
DT 01-OCT-1996 (Rel. 34, Created)  
DT 01-OCT-1996 (Rel. 34, Last sequence update)  
DT 10-OCT-2003 (Rel. 42, Last annotation update)  
DE Adenylosuccinate synthetase (EC 6.3.4.4) (IMP--aspartate ligase)  
DE (AdSS) (AMPase).  
GN PURA  
OS Thiobacillus ferrooxidans.  
OC Bacteria; Proteobacteria; Gammaproteobacteria; Acidithiobacillales;  
OC Acidithiobacillaceae; Acidithiobacillus.  
OX NCBI\_TaxID=920;  
RN [1]  
RP. SEQUENCE FROM N.A.  
RC STRAIN=Rel;  
RX MEDLINE=92283830; PubMed=1317860;  
RA Kusano T., Takeshima T., Sugawara K., Inoue C., Shiratori T., Yano T.,  
RA Fukumori Y., Yamanka T.;  
RT "Molecular cloning of the gene encoding Thiobacillus ferrooxidans  
RT Fe(II) oxidase. High homology of the gene product with HiFIP.";  
RL J. Biol. Chem. 267:11242-11247(1992).  
CC -1- FUNCTION: Plays an important role in the de novo pathway of purine  
CC nucleotide biosynthesis.  
CC -1- CATALYTIC ACTIVITY: GTP + IMP + L-aspartate = GDP + phosphate +  
CC adenylosuccinate.  
CC -1- COFACTOR: Binds 1 magnesium ion per subunit (By similarity).  
CC -1- PATHWAY: AMP biosynthesis; first committed step.  
CC -1- SUBUNIT: Homodimer (By similarity).  
CC -1- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).  
CC -1- SIMILARITY: Belongs to the adenylosuccinate synthetase family.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; X57324; CAA0593.1; -;  
CC PIR; S23258; S23258.  
CC HSSP; P12283; IADE.  
CC HAMAP; MF 00011; -;  
CC InterPro; IPR001114; Asucc synthetase.  
CC Pfam; PF00709; adenylosucc synt; 1.  
CC ProDom; PD001188; Asucc synthetase; 1.  
CC TIGRFAMs; TIGR00184; pura; 1.  
CC PROSITE; PS00513; ADENYLOSUCCIN SYN\_2; 1.  
CC PROSITE; PS01266; ADENYLOSUCCIN SYN\_1; 1.  
CC Purine biosynthesis; Ligase; GTP-binding; Metal-binding; Magnesium.  
FT NP BIND 13 19  
FT ACT\_SITE 141 141 BY SIMILARITY.  
FT ACT\_SITE 148 148 BY SIMILARITY.  
FT METAL 14 14 MAGNESIUM (BY SIMILARITY).  
FT METAL 41 41 MAGNESIUM (VIA CARBONYL OXYGEN) (BY  
FT SIMILARITY).  
SQ SEQUENCE 429 AA; 46355 MW; CSB2F39C3E105719 CRC64;  
  
Query Match 3.3%; Score 8; DB 1; Length 429;  
Best Local Similarity 100.0%; Pred. No. 3.4;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
Qy 148 LGYVLGIT 155  
Db 260 LGYVLGIT 267  
|||||||  
|||  
  
RESULT 2  
HGFL MOUSE  
ID HGFL MOUSE STANDARD; PRT; 716 AA.  
AC P26928;  
DT 01-AUG-1992 (Rel. 23, Created)

DT 01-AUG-1992 (Rel. 23, Last sequence update)  
DT 10-OCT-2003 (Rel. 42, Last annotation update)  
DE Hepatocyte growth factor-like protein precursor (Macrophage  
DE stimulatory protein) (MSP).  
GN MST1 OR HGFL  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP. SEQUENCE FROM N.A.  
RC STRAIN=BALB/C; TISSUE=Liver;  
RX MEDLINE=9202017; PubMed=1832957;  
RA Friesner Degen S.J., Stuart L.A., Han S., Jamison C.S.;  
RT "Characterization of the mouse CDNA and gene coding for a hepatocyte  
RT growth factor-like protein: expression during development.";  
RL Biochemistry 30:9781-9791(1991).  
CC -1- FUNCTION: Probably has no proteolytic activity, since crucial AA  
CC characteristic of serine proteases catalytic sites are not  
CC conserved.  
CC -1- TISSUE SPECIFICITY: Liver. Lower levels in lung, placenta and  
CC adrenal.  
CC -1- DEVELOPMENTAL STAGE: Is expressed at low levels during gestation.  
CC Just before birth the level increases dramatically and remains  
CC stable afterwards.  
CC -1- PTM: MAY BE CLEAVED AFTER AA 488, TO YIELD A TWO-CHAIN MOLECULE  
CC HELD TOGETHER BY DISULFIDE BONDS, OR CLEAVED INTO TWO SEPARATE  
CC POLYPEPTIDES.  
CC -1- SIMILARITY: Belongs to peptidase family S1. Plasminogen subfamily.  
CC -1- SIMILARITY: Contains 4 Kringle domains.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; M74180; AAA50166.1; -;  
CC PIR; M74181; AAA50167.1; -;  
CC PIR; A40332; A40332.  
CC HSSP; P00747; IKEN.  
CC MEROPS; S01.975; -;  
CC MGD; MGI:96080; Mst1.  
CC GO; GO:0007566; P:embryo implantation; IC.  
CC InterPro; IPR009003; Cys Ser trypsin.  
CC InterPro; IPR000001; Kringle.  
CC InterPro; IPR003014; PAN.  
CC InterPro; IPR003609; Pan app.  
CC InterPro; IPR001254; Peptidase\_S1.  
CC InterPro; IPR001314; Peptidase\_S1A.  
CC InterPro; IPR003966; Peptidase\_S1A\_pr.  
CC Pfam; PF00051; kringle; 4.  
CC Pfam; PF00024; PAN; 1.  
CC Pfam; PF00089; trypsin; 1.  
CC PRINTS; PR00722; CHYMOTRYPSIN.  
CC PRINTS; PR00018; KRINGLE.  
CC PRINTS; PR01505; PROTHROMBIN.  
CC ProDom; PD000395; Kringle; 4.  
CC SMART; SM00130; KR; 4.  
CC SMART; SM00473; PAN AP; 1.  
CC SMART; SM00020; TRYD\_SPC; 1.  
CC PROSITE; PS00021; KRINGLE\_1; 4.  
CC PROSITE; PS00070; KRINGLE\_2; 4.  
CC PROSITE; PS00240; TRIPSN\_DOM; 1.  
CC Kringle; Glycoprotein; Serine protease homolog; Repeat; Signal.  
FT SIGNAL 1 18  
FT CHAIN 19 716  
FT DOMAIN 19 109  
FT DOMAIN 110 186  
FT DOMAIN 191 268  
FT DOMAIN 292 370  
FT KRINGLE 1.  
FT KRINGLE 2.  
FT KRINGLE 3.

```

FT DOMAIN 379 457 KRINGLE 4.
FT FT 489 716 SERINE PROTEASE-LIKE.
FT DISULFID 56 78 BY SIMILARITY.
FT DISULFID 60 86 BY SIMILARITY.
FT DISULFID 110 186 BY SIMILARITY.
FT DISULFID 131 169 BY SIMILARITY.
FT DISULFID 157 181 BY SIMILARITY.
FT DISULFID 191 268 BY SIMILARITY.
FT DISULFID 194 333 INTERCHAIN (BY SIMILARITY).
FT DISULFID 212 281 BY SIMILARITY.
FT DISULFID 240 283 BY SIMILARITY.
FT DISULFID 292 370 BY SIMILARITY.
FT DISULFID 313 352 BY SIMILARITY.
FT DISULFID 341 364 BY SIMILARITY.
FT DISULFID 379 457 BY SIMILARITY.
FT DISULFID 400 440 BY SIMILARITY.
FT DISULFID 428 452 BY SIMILARITY.
FT DISULFID 477 533 INTERCHAIN (BY SIMILARITY).
FT DISULFID 512 528 BY SIMILARITY.
FT DISULFID 607 672 BY SIMILARITY.
FT DISULFID 637 651 BY SIMILARITY.
FT DISULFID 662 690 BY SIMILARITY.
FT CARBOHYD 72 72 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 173 173 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 305 305 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 620 620 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CONFLICT 19 19 P -> Q (IN GENOMIC SEQUENCE).
SQ SEQUENCE 716 AA; 80588 MW; BICE02EF85213ACC CRC64;

Query Match 3.3%; Score 8; DB 1; Length 716;
Best Local Similarity 100.0%; Pred. No. 5,3;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 55 DRPGWCY 62
DB 163 DRPGWCY 170

RESULT 3
PCRA_MYCTU STANDARD; PRT; 771 AA.
ID PCRA_MYCTU STANDARD; PRT; 771 AA.
AC P71561;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DE 10-OCT-2003 (Rel. 42, Last annotation update)
DE ATP-dependent DNA helicase PCRA (EC 3.6.1.-)
GN PCRA OR IVD OR RV0949 OR MT0976 OR MTCY10D7.25C OR MB0974.
OS Mycobacterium tuberculosis, and
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773, 1765;
RN [1]
RP SEQUENCE FROM N.A.
RC SPECIES=M.tuberculosis; STRAIN=H37RV;
RX MEDLINE=98295987; PubMed=9634230;
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekala F.,
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
RA Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;
RT "Deciphering the biology of Mycobacterium tuberculosis from the
RT complete genome sequence.";
RL Nature 393:537-544 (1998).
RN [2]
RP SEQUENCE FROM N.A.
RC SPECIES=M.tuberculosis; STRAIN=CDC 1551 / Oshkosh;
RX MEDLINE=22206494; PubMed=12218036;
RA Fleischnann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA Peterson J., DeBoy R., Dodson R., Gwinn M., Haft D., Hickey B.,

```

```

RA Kelonay J.F., Nelson W.C., Umavay L.A., Ermolaeva M., Salzberg S.L.,
RA Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
RA Bishai W., Jacobs W.R. Jr., Venter J.C., Fraser C.M.;
RT "Whole-genome comparison of Mycobacterium tuberculosis clinical and
RT laboratory strains.";
RN [3]
RP J. Bacteriol. 184:5479-5490 (2002).
RP SEQUENCE FROM N.A.
RC SPECIES=M.bovis; STRAIN=AF2122/97;
RX MEDLINE=22709107; PubMed=12788972;
RA Garnier T., Eiglmeier K., Camus J.-C., Medina N., Mansoor H.,
RA Pryor M., Duthoy S., Grondin S., Lacroix C., Monsemp C., Simon S.,
RA Harris B., Atkin R., Doggett J., Mayes R., Keating L., Wheeler P.R.,
RA Parkhill J., Barrell B.G., Cole S.T., Gordon S.V., Wheeler R.G.;
RT "The complete genome sequence of Mycobacterium bovis.";
RL Proc. Natl. Acad. Sci. U.S.A. 100:7877-7882 (2003).
CC -!- SIMILARITY: Belongs to the helicase family. UvrD subfamily.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC -----
CC EMBL; Z79700; CAB02001.1; -.
CC EMBL; AE006982; AAK45224.1; -.
CC EMBL; BX248337; CAD93835.1; -.
CC PIR; C70716; C70716.
CC HSP; P56255; 1PJR.
CC TIGR; MT0976; -.
CC TuberculList; RV0949; -.
CC InterPro; IPR005751; PcrA.
CC InterPro; IPR000212; UvrD-helicase.
CC Pfam; PF00580; UvrD-helicase; 1.
CC TIGRFAMs; TIGR01073; pcrA; 1.
KW DNA-binding; Hydrolase; Helicase; ATP-binding; DNA repair;
KW Complete Proteome.
FT NP_BIND 42 49 ATP (POTENTIAL).
SQ SEQUENCE 771 AA; 85049 MW; DBAA4E151F4E2C9 CRC64;

Query Match 3.3%; Score 8; DB 1; Length 771;
Best Local Similarity 100.0%; Pred. No. 5,6;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 37 SAPVSGAG 44
DB 692 SAPVSGAG 699

RESULT 4
CH10_AMOPS STANDARD; PRT; 96 AA.
ID CH10_AMOPS STANDARD; PRT; 96 AA.
AC P26005;
DT 01-MAY-1992 (Rel. 22, Created)
DT 01-MAY-1992 (Rel. 22, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE 10 kDa chaperonin (protein Cpn10) (groES protein).
GN GROS OR GROS.
OS Amoeba proteus symbiotic bacterium.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Legionellales;
OC Legionellaceae.
OX NCBI_TaxID=2728;
RN [1]
RP SEQUENCE FROM N.A.
RA Ahn T.I., Leeu H.K., Kwak I.H., Jeon K.W.;
RT "Nucleotide sequence and temperature-dependent expression of groEL
RT gene isolated from symbiotic bacteria of Amoeba proteus.";
RL Endocyt. Cell Res. 8:33-44 (1991).
CC -!- FUNCTION: Binds to Cpn60 in the presence of Mg-ATP and suppresses
CC the ATPase activity of the latter.
CC -!- SUBUNIT: Heptamer of 7 subunits arranged in a ring (By

```

```

CC CC similarity).
CC CC -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC CC -!- SIMILARITY: Belongs to the groES chaperonin family.
CC CC
CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC CC the European Bioinformatics Institute. There are no restrictions on its
CC CC use by non-profit institutions as long as its content is in no way
CC CC modified and this statement is not removed. Usage by and for commercial
CC CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC CC or send an email to license@isb-sib.ch).
CC CC
CC CC EMBL; M86549; AAC09380.1; -.
CC CC PIR; JC2561; JC2561.
CC CC HAMAP; MF_00580; -. 1.
CC CC InterPro; IPR001476; Chaperin_Cpn10.
CC CC Pfam; PF00166; cpn10; 1.
CC CC PRINTS; PR00297; CHAPERONIN10.
CC CC ProDom; PD000566; Chaperin_Cpn10; 1.
CC CC PROSITE; PS00681; CHAPERONINS_CPN10; 1.
CC CC Chaperone.
CC CC
CC CC SEQUENCE 96 AA; 10618 MW; F6D5E1A8056B899B CRC64;
CC CC
CC CC Query Match 2.9%; Score 7; DB 1; Length 96;
CC CC Best Local Similarity 100.0%; Pred. No. 10;
CC CC Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
CC CC
CC CC QY 160 IIAIGAG 166
CC CC |||||
CC CC 40 IIAIGAG 46
CC CC
CC CC RESULT 5
CC CC YL66_ARCFU STANDARD; PRT; 154 AA.
CC CC ID YL66_ARCFU
CC CC AC C28116;
CC CC DT 16-OCT-2001 (Rel. 40, Created)
CC CC DT 16-OCT-2001 (Rel. 40, Last sequence update)
CC CC DT 16-OCT-2001 (Rel. 40, Last annotation update)
CC CC DE Hypothetical protein AF2166.
CC CC GN AF2166.
CC CC OS Archaeoglobus fulgidus.
CC CC OC Archaea; Euryarchaeota; Archaeoglobi; Archaeoglobales;
CC CC OC Archaeoglobaceae; Archaeoglobus.
CC CC OX NCBI_TaxID=2234;
CC CC RN [1]
CC CC RP SEQUENCE FROM N.A.
CC CC RC STEAIN=VC-16 / DSM 4304 / ATCC 49558;
CC CC RX MEDLINE=98049343; Pubmed=9382475;
CC CC RA Klenk H.-P., Clayton R.A., Tomb J.-F., White O., Nelson K.E.,
CC CC Ketchum K.A., Dodson R.J., Winn M., Hickey E.K., Peterson J.D.,
CC CC Richardson D.L., Kervagge A.R., Graham D.E., Kyprides N.C.,
CC CC Fleischmann R.D., Quackenbush J., Lee N.H., Sutton G.G., Gill S.,
CC CC Kirkness E.F., Dougherty B.A., McKenney K., Adams M.D., Loftus B.,
CC CC Peterson S., Reich C.I., McNeil L.K., Badger J.H., Glodek A., Zhou L.,
CC CC Overbeek R., Goynne J.D., Weidman J.F., McDonald L., Utterback T.,
CC CC Cotton M.D., Spriggs T., Artach P., Kaine B.P., Sykes S.M.,
CC CC Sadow P.W., D'Andrea K.P., Bowman C., Fujii C., Garland S.A.,
CC CC Mason T.M., Olsen G.J., Fraser C.M., Smith H.O., Woese C.R.,
CC CC Venter J.C.;
CC CC "The complete genome sequence of the hyperthermophilic, sulphate-
CC CC reducing archaeon Archaeoglobus fulgidus.";
CC CC Nature 390:364-370(1997).
CC CC -!- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
CC CC
CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC CC the European Bioinformatics Institute. There are no restrictions on its
CC CC use by non-profit institutions as long as its content is in no way
CC CC modified and this statement is not removed. Usage by and for commercial
CC CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC CC or send an email to license@isb-sib.ch).
CC CC

```

```

CC CC EMBL; AB000955; AAB89089.1; -.
CC CC PIR; F69520; F69520.
CC CC TIGR; AF2166; -.
CC CC KW Hypothetical protein; Transmembrane; Complete proteome.
CC CC FT TRANSMEM 15 37 POTENTIAL.
CC CC FT TRANSMEM 58 80 POTENTIAL.
CC CC FT TRANSMEM 95 116 POTENTIAL.
CC CC FT TRANSMEM 123 145 POTENTIAL.
CC CC SQ SEQUENCE 154 AA; 17266 MW; 1CE41361CD8FA468 CRC64;
CC CC
CC CC Query Match 2.9%; Score 7; DB 1; Length 154;
CC CC Best Local Similarity 100.0%; Pred. No. 15;
CC CC Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
CC CC
CC CC QY 166 GIILGYS 172
CC CC |||||
CC CC 104 GIILGYS 110
CC CC
CC CC RESULT 6
CC CC ST04_CAEEL STANDARD; PRT; 281 AA.
CC CC ID ST04_CAEEL
CC CC AC Q22165; Q9N644;
CC CC DT 01-NOV-1997 (Rel. 35, Created)
CC CC DT 28-FEB-2003 (Rel. 41, Last sequence update)
CC CC DT 28-FEB-2003 (Rel. 41, Last annotation update)
CC CC DE Stomatin protein 4.
CC CC GN STO-4 OR Y71H9A.3/TO4P8.5.
CC CC OS Caenorhabditis elegans.
CC CC OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
CC CC OC Rhabditidae; Peloderinae; Caenorhabditis.
CC CC OX NCBI_TaxID=6239;
CC CC RN [1]
CC CC RP SEQUENCE FROM N.A.
CC CC RC STRAIN=Bristol N2;
CC CC RA Lennard N., White S.;
CC CC RL Submitted (NOV-1995) to the EMBL/GenBank/DBJ databases.
CC CC RN [2]
CC CC RP REVISIONS.
CC CC RA Durbin R.;
CC CC RL Submitted (AUG-1998) to the EMBL/GenBank/DBJ databases.
CC CC -!- SIMILARITY: Belongs to the band 7 / mec-2 family.
CC CC
CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC CC the European Bioinformatics Institute. There are no restrictions on its
CC CC use by non-profit institutions as long as its content is in no way
CC CC modified and this statement is not removed. Usage by and for commercial
CC CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC CC or send an email to license@isb-sib.ch).
CC CC
CC CC EMBL; AL032659; CAB82215.1; -.
CC CC EMBL; Z66565; CAB82215.1; JOINED.
CC CC EMBL; Z66565; CAB76415.1; -.
CC CC EMBL; AL032659; CAB76415.1; JOINED.
CC CC WormPep; Y71H9A.3; CE22957.
CC CC InterPro; IPR001107; Band 7.
CC CC InterPro; IPR001972; Stomatin.
CC CC Pfam; PF01145; Band 7; 1.
CC CC PRINTS; PR00721; STOMATIN.
CC CC SMART; SM00244; PHB; 1.
CC CC PROSITE; PS01270; BAND_7; 1.
CC CC KW Transmembrane; Multigene family.
CC CC FT TRANSMEM 28 48 POTENTIAL.
CC CC FT DOMAIN 202 238 ALA-RICH.
CC CC SQ SEQUENCE 281 AA; 30843 MW; D56648D4682FC9C CRC64;
CC CC
CC CC Query Match 2.9%; Score 7; DB 1; Length 281;
CC CC Best Local Similarity 100.0%; Pred. No. 25;
CC CC Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
CC CC
CC CC QY 196 TLPLSAF 202

```

Db	41	TUPLSAF	47	
RESULT 7				
SC14_CANAL		STANDARD;	PRT;	301 AA.
ID	SC14_CANAL			
AC	P46250;			
DT	01-NOV-1995 (Rel. 32, Created)			
DT	01-NOV-1995 (Rel. 32, Last sequence update)			
DT	28-FEB-2003 (Rel. 41, Last annotation update)			
DE	SEC14 cytosolic factor (Phosphatidylinositol/phosphatidylcholine transfer protein) (PI/PC TP).			
GN	SEC14.			
OS	Candida albicans (Yeast).			
OC	Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;			
OC	Saccharomycetales; mitosporic Saccharomycetales; Candida.			
OX	NCBI_TaxID=5476;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=ATCC 64385 / 1001;			
RC	MEDLINE=97051600; PubMed=8896277;			
RA	Montecoliva L., Sanchez M., Pla J., Gil C., Nombela C.;			
RT	"Cloning of Candida albicans SEC14 gene homologue coding for a putative essential function.";			
RL	Yeast 12:1097-1105 (1996).			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=SGY243;			
RC	MEDLINE=98048477; PubMed=9387231;			
RA	Riggie P.J., Slobodkin I.V., Brown D.H. Jr., Hanson M.P.,			
RA	Volker T.L., Kumamoto C.A.;			
RT	"Two transcripts, differing at their 3' ends, are produced from the Candida albicans SEC14 gene.";			
RL	Microbiology 143:3527-3535 (1997).			
CC	-1- FUNCTION: Required for transport of secretory proteins from the Golgi complex. Catalyzes the transfer of phosphatidylinositol and phosphatidylcholine between membranes in vitro (By similarity).			
CC	-1- SUBCELLULAR LOCATION: Associated with the Golgi complex as a peripheral membrane protein (By similarity).			
CC	-1- SIMILARITY: Contains 1 CRAL-TRIO domain.			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a> or send an email to <a href="mailto:license@isb-sib.ch">license@isb-sib.ch</a> ).			
DR	EMBL; X81937; CAA57490.1; -			
DR	EMBL; U61975; ABA41491.1; -			
DR	HSP; P24280; 1AUA.			
DR	InterPro; IPR001251; CRAL_TRIO_C.			
DR	InterPro; IPR008273; CRAL_TRIO_N.			
DR	InterPro; IPR01071; RetBind/tocTrans.			
DR	Pfam; PF00650; CRAL_TRIO; 1.			
DR	Pfam; PF03765; CRAL_TRIO_N; 1.			
DR	PRINTS; PR00180; CRÉTINALDHP.			
DR	SMART; SM00516; SEC14; 1.			
DR	PROSITE; PS00191; CRAL_TRIO; 1.			
DR	PROSITE; PS00191; CRAL_TRIO; 1.			
KW	Transport; Protein transport; Golgi stack.			
FT	DOMAIN 99 CRAL-TRIO.			
SQ	SEQUENCE 301 AA; 34709 MW; 1F194A8EA8B525BE CRC64;			
Query Match	2.9%; Score 7; DB 1; Length 301;			
Best Local Similarity	100.0%; Pred. No. 27;			
Matches	7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;			
QY	168 ILGYSYK 174			
Db	242 ILGYSYK 248			
RESULT 8				
GPDA_HABIN		STANDARD;	PRT;	335 AA.
ID	GPDA_HABIN			
AC	P43798;			
DT	01-NOV-1995 (Rel. 32, Created)			
DT	01-NOV-1995 (Rel. 32, Last sequence update)			
DT	28-FEB-2003 (Rel. 41, Last annotation update)			
DE	Glycerol-3-phosphate dehydrogenase [NAD(P)+] (EC 1.1.1.94) (NAD(P)H-dependent glycerol-3-phosphate dehydrogenase).			
GN	GP8A OR H10605.			
OS	Haemophilus influenzae.			
OC	Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;			
OC	Pasteurellaceae; Haemophilus.			
OX	NCBI_TaxID=727;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=RD / KW20 / ATCC 51907;			
RC	MEDLINE=95350630; PubMed=7542800;			
RA	Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,			
RA	Kerlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,			
RA	McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,			
RA	Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M.,			
RA	Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,			
RA	Utterback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,			
RA	Fine L.D., Fritchman J.L., Fuhrmann J.L., Geoghagen N.S.M.,			
RA	Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,			
RA	Venter J.C.;			
RT	"Whole-genome random sequencing and assembly of Haemophilus influenzae Rd.";			
RL	Science 269:496-512 (1995).			
CC	-1- CATALYTIC ACTIVITY: Sn-glycerol 3-phosphate + NAD(P) (+) =			
CC	glycerone phosphate + NAD(P)H.			
CC	-1- PATHWAY: De novo phospholipid biosynthesis; glycerol-3 phosphate formation.			
CC	-1- SUBCELLULAR LOCATION: Cytoplasmic (Probable).			
CC	-1- SIMILARITY: Belongs to the NAD-dependent glycerol-3-phosphate dehydrogenase family.			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a> or send an email to <a href="mailto:license@isb-sib.ch">license@isb-sib.ch</a> ).			
DR	EMBL; U32743; AAC22264.1; -			
DR	PIR; F64080; F64080.			
DR	TIGR; H10605; -			
DR	HAMAP; MF_00394; -; 1.			
DR	InterPro; IPR008927; 6GDH_C-like.			
DR	InterPro; IPR008168; NAD_Gly3P_dh.			
DR	InterPro; IPR006109; NAD_Gly3P_dom.			
DR	Pfam; PF01210; NAD_Gly3P_dh; 1.			
DR	PRINTS; PR00077; GPDHGRNASE.			
DR	PRODOM; PD001278; NAD_Gly3P_dom; 1.			
DR	PROSITE; PS00957; NAD_G3PDH; 1.			
KW	Phospholipid biosynthesis; Oxidoreductase; NAD; Complete proteome.			
SQ	SEQUENCE 335 AA; 36394 MW; 4913F8F2203B7CFC CRC64;			
Query Match	2.9%; Score 7; DB 1; Length 335;			
Best Local Similarity	100.0%; Pred. No. 29;			
Matches	7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;			
QY	161 IAIAGGI 167			
Db	198 IAIAGGI 204			
RESULT 9				
Y780_METJA		STANDARD;	PRT;	335 AA.
ID	Y780_METJA			

Q58190;  
01-NOV-1997 (Rel. 35, Created)  
01-NOV-1997 (Rel. 35, Last sequence update)  
16-OCT-2001 (Rel. 40, Last annotation update)  
Hypothetical protein MJ0780.  
GN MJ0780  
OS Methanococcus jannaschii.  
CC Archaea; Euryarchaeota; Methanococci; Methanococcales;  
CC Methanocaldococcaceae; Methanocaldococcus.  
CC NCBI\_TaxID=2190;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=JAL-1 / DSM 2661 / ATCC 43067;  
RX MEDLINE=96337999; PubMed=9688087;  
RA Bult C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,  
RA Sutton G.G., Blake J.A., FitzGerald L.M., Clayton R.A., Gocayne J.D.,  
RA Kerlavage A.R., Dougherty B.A., Tomb J.-F., Adams M.D., Reich C.L.,  
RA Overbeek R., Kirkness E.F., Weinstock K.G., Merrick J.M., Glodek A.,  
RA Scott J.L., Geoghegan N.S.M., Weidman J.F., Fuhrmann J.L., Nguyen D.,  
RA Utterback T.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,  
RA Cotton M.D., Roberts K.M., Hurst M.A., Kaine B.P., Borodovsky M.,  
RA Klenk H.-P., Fraser C.M., Smith H.O., Woese C.R., Venter J.C.;  
RA "Complete genome sequence of the methanogenic archaeon, Methanococcus  
jannaschii.";  
RN Science 273:1058-1073(1996).  
CC -!- SUBCELLULAR LOCATION: Integral membrane protein (Potential).  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; U67522; AAB98779.1; -  
DR PIR; D64397; D64397.  
DR TIGR; MJ0780; -  
DR InterPro; IPR002814; Flaj.  
DR Pfam; PF01961; Flaj arch: 1.  
KW Hypothetical protein, Transmembrane; Complete proteome.  
FT TRANSMEM 104 124 POTENTIAL.  
FT TRANSMEM 128 148 POTENTIAL.  
FT TRANSMEM 280 300 POTENTIAL.  
FT TRANSMEM 310 330 POTENTIAL.  
SQ SEQUENCE 335 AA; 38138 MW; C6028EDBA1455C08 CRC64;  
  
Query Match 2.9%; Score 7; DB 1; Length 335;  
Best Local Similarity 100.0%; Pred. No. 29;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 163 IGAGIIL 169  
Db 137 IGAGIIL 143  
  
RESULT 10  
GPDA\_HAEU STANDARD; PRT; 336 AA.  
AC Q7VNA0;  
DT 15-MAR-2004 (Rel. 43, Created)  
DT 15-MAR-2004 (Rel. 43, Last sequence update)  
DT 15-MAR-2004 (Rel. 43, Last annotation update)  
DE Glycerol-3-phosphate dehydrogenase [NAD(P)+] (EC 1.1.1.94) (NAD(P)H-  
DE dependent glycerol-3-phosphate dehydrogenase).  
GN GPDA OR HD660.  
OS Haemophilus ducreyi.  
CC Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;  
CC Pasteurellaceae; Haemophilus.  
CC NCBI\_TaxID=730;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=35000HP / ATCC 700724;

Munson R.S. Jr., Ray W.C., Mahairas G., Sabo P., Mungur R.,  
Johnson L., Nguyen D., Wang J., Forst C., Hood L.;  
"The complete genome sequence of Haemophilus ducreyi.";  
Submitted (JUN-2003) to the EMBL/GenBank/DBJ databases.  
-!- CATALYTIC ACTIVITY: Sn-glycerol 3-phosphate + NAD(P)+ =  
CC glycerone phosphate + NAD(P)H.  
CC -!- PATHWAY: De novo phospholipid biosynthesis; glycerol-3 phosphate  
CC formation.  
CC -!- SUBCELLULAR LOCATION: Cytoplasmic (Probable).  
CC -!- SIMILARITY: Belongs to the NAD-dependent glycerol-3-phosphate  
CC dehydrogenase family.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; AE017152; AAP95585.1; -  
DR HMAP; MF 00394; -1  
DR InterPro; IPR006168; NAD\_Gly3P\_dh.  
DR InterPro; IPR006109; NAD\_Gly3P\_dom.  
DR Pfam; PF01210; NAD\_Gly3P\_dh; 1.  
DR ProDom; PD001278; NAD\_Gly3P\_dom; 1.  
DR PROSITE; PS00957; NAD\_G3PDH; 1.  
KW Phospholipid biosynthesis; Oxidoreductase; NAD; Complete proteome.  
SQ SEQUENCE 336 AA; 36575 MW; 70D4E2D2969136C4 CRC64;  
  
Query Match 2.9%; Score 7; DB 1; Length 336;  
Best Local Similarity 100.0%; Pred. No. 30;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 161 IAIGAGI 167  
Db 199 IAIGAGI 205  
  
RESULT 11  
XYLR\_LACPE STANDARD; PRT; 388 AA.  
AC P21940;  
DT 01-MAY-1991 (Rel. 18, Created)  
DT 01-MAR-1992 (Rel. 21, Last sequence update)  
DT 01-NOV-1995 (Rel. 32, Last annotation update)  
DE Xylose repressor.  
GN XYLR.  
OS Lactobacillus pentosus.  
CC Bacteria; Firmicutes; Lactobacillales; Lactobacillaceae;  
CC Lactobacillus.  
CC NCBI\_TaxID=1569;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=MD353.  
RX MEDLINE=92079891; PubMed=1660563;  
RA Lokman B.C., van Santen P., Verdoes J.C., Kruse J., Leer R.J.,  
RA Poseno M., Pouwels F.H.;  
RA "Organization and characterization of three genes involved in  
RT D-xylose catabolism in Lactobacillus pentosus.";  
RL Mol. Gen. Genet. 230:161-169(1991).  
CC -!- FUNCTION: TRANSCRIPTIONAL REPRESSOR OF XYLOSE-UTILIZING ENZYMES.  
CC -!- SIMILARITY: BELONGS TO THE ROK (NAGC/XYLR) FAMILY.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; M57384; AAA25257.1; -

DR PIR; S18560; S18560.  
DR InterPro; IPR004839; NHtransf\_1\_BS.  
DR Pfam; PF00155; aminotran\_1\_2; 1.  
DR PRINTS; PR00799; TRANSAMINASE.  
DR PROSITE; PS00105; AA TRANSFER CLASS 1; 1.  
KW Transcription regulation; DNA-binding; Repressor; Xylose metabolism.  
FT H-T-H MOTIF (BY SIMILARITY).  
SQ DNA BIND 31 50  
SEQUENCE 388 AA; 43604 MW; 2CB03049F6A75412 CRC64;  
Query Match 2.9%; Score 7; DB 1; Length 388;  
Best Local Similarity 100.0%; Pred. No. 33;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 163 IGAGIIL 169  
DB 219 IGAGIIL 225  
RESULT 12  
AAAT\_PSEAE STANDARD; PRT; 398 AA.  
AC P72173;  
DT 01-NOV-1997 (Rel. 35, Created)  
DT 16-OCT-2001 (Rel. 40, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE Aspartate aminotransferase (EC 2.6.1.1) (Transaminase A) (AspAT).  
GN ASFC OR PA3139.  
OS Pseudomonas aeruginosa.  
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;  
OC Pseudomonadaceae; Pseudomonas.  
OX NCBI\_TaxID=287;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=ATCC 15692 / PA01;  
RA MEDLINE=20437337; PubMed=10984043;  
RA Stover C.K., Pham X.-Q.T., Erwin A.L., Mizoguchi S.D., Warren P.,  
RA Hickey M.J., Brinkman P.S.L., Hufnagle W.O., Kowalik D.J., Lagrou M.,  
RA Garber R.L., Goltzy L., Tolentino E., Westbrook-Wadman S., Yuan Y.,  
RA Brody L.L., Coulter S.N., Folger K.R., Kas A., Larbig K., Lim R.M.,  
RA Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,  
RA Reizer J., Saier M.H., Hancock R.E.W., Lory S., Olson M.V.;  
RT "Complete genome sequence of Pseudomonas aeruginosa PA01, an  
RT opportunistic pathogen.";  
RL Nature 406:959-964(2000).  
[2]  
RP SEQUENCE OF 1-75 FROM N.A.  
RC STRAIN=ATCC 15692 / PA01;  
RX MEDLINE=96404811; PubMed=9808952;  
RA Rivera E., Vila L., Barbe J.;  
RT "The *uvrB* gene of Pseudomonas aeruginosa is not DNA damage  
RT inducible.";  
RL J. Bacteriol. 178:5550-5554(1996).  
CC -!- CATALYTIC ACTIVITY: L-aspartate + 2-oxoglutarate = oxaloacetate +  
CC L-glutamate.  
CC -!- COFACTOR: Pyridoxal phosphate (By similarity).  
CC -!- SUBUNIT: Homodimer (By similarity).  
CC -!- SUBCELLULAR LOCATION: Cytoplasmic.  
CC -!- SIMILARITY: Belongs to class-I of pyridoxal-phosphate-dependent  
CC aminotransferases.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; AE004738; AAG06527.1; -.  
CC EMBL; X93486; CA63758.1; -.  
CC F01; B3252; B83252.  
CC HSP; P04693; 3TAT.  
CC InterPro; IPR004839; Aminotrans\_1/II.  
CC InterPro; IPR000796; Asp\_trans.

DR PIR; S18560; S18560.  
DR InterPro; IPR000600; ROK.  
DR Pfam; PF00480; ROK; 1.  
DR PROSITE; PS01125; ROK; 1.  
KW Transcription regulation; DNA-binding; Repressor; Xylose metabolism.  
FT H-T-H MOTIF (BY SIMILARITY).  
SQ DNA BIND 31 50  
SEQUENCE 388 AA; 43604 MW; 2CB03049F6A75412 CRC64;  
Query Match 2.9%; Score 7; DB 1; Length 388;  
Best Local Similarity 100.0%; Pred. No. 33;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 163 IGAGIIL 169  
DB 219 IGAGIIL 225  
RESULT 13  
PURA\_BRUAB STANDARD; PRT; 402 AA.  
AC P52004;  
DT 01-OCT-1996 (Rel. 34, Created)  
DT 30-MAY-2000 (Rel. 39, Last sequence update)  
DT 10-OCT-2003 (Rel. 42, Last annotation update)  
DE Adenylosuccinate synthetase (EC 6.3.4.4) (IMP--aspartate ligase)  
DE (AGSS) (AMPSase) (Fragment).  
GN PURA.  
OS Brucella abortus.  
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;  
OC Brucellaceae; Brucella.  
OX NCBI\_TaxID=235;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=2308;  
RA Tatum F.M., Steckelberg M.A.;  
RL Submitted (SEP-1995) to the EMBL/GenBank/DBJ databases.  
[2]  
RP IDENTIFICATION OF PROBABLE C-TERMINAL FRAMESHIFT.  
RA Michoud K.;  
RL Unpublished observations (NOV-1999).  
CC -!- FUNCTION: Plays an important role in the de novo pathway of purine  
CC nucleotide biosynthesis.  
CC -!- CATALYTIC ACTIVITY: GTP + IMP + L-aspartate = GDP + phosphate +  
CC adenylosuccinate.  
CC -!- COFACTOR: Binds 1 magnesium ion per subunit (By similarity).  
CC -!- PATHWAY: AMP biosynthesis; first committed step.  
CC -!- SUBUNIT: Homodimer (By similarity).  
CC -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).  
CC -!- SIMILARITY: Belongs to the adenylosuccinate synthetase family.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; L43054; AA475455.1; -.  
CC HSP; P12283; IADE.  
CC F01; MF\_00011; -; 1.  
CC InterPro; IPR001114; Asucc synthetase.  
CC Pfam; PF00709; Adenylysucc synt; 1.  
CC ProDom; PD001188; Asucc synthase; 1.  
CC TIGRFAM; TIGR00184; PURA; 1.  
CC PROSITE; PS00513; ADENYLOSUCCIN SYN 2; 1.  
CC PROSITE; PS01266; ADENYLOSUCCIN SYN 1; 1.  
KW Purine biosynthesis; Ligase; GTP-binding; Metal-binding; Magnesium.  
FT NP BIND 12 18  
FT ACT\_SITE 140 140  
FT ACT\_SITE 147 147  
FT ACT\_SITE 13 13  
FT METAL 40 40  
FT METAL

(BY  
MAGNESIUM (VIA CARBONYL OXYGEN) (BY

```
FT CONFLICT 393 402 SIMILARITY).
FT NON TER 402 402
SQ SEQUENCE 402 AA; 43919 MW; F443CFAE8D10BE75 CRC64;

Query Match 2.9%; Score 7; DB 1; Length 402;
Best Local Similarity 100.0%; Pred. No. 34;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 149 GYVLGIT 155
    |||||
Db 258 GYVLGIT 264

RESULT 14
PURA BRUME STANDARD; PRT; 429 AA.
ID PURA BRUME STANDARD; PRT; 429 AA.
AC QYVILU; Q8P220;
DT 10-OCT-2003 (Rel. 42, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Adenylosuccinate synthetase (EC 6.3.4.4) (IMP--aspartate ligase)
DE (AdSS) (AMPSase).
GN PURA OR BMEI0351 OR BR1683.
OS Brucella melitensis, and
OS Brucella suis
OS Brucella melitensis, and
OS Brucella suis
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC Brucellaceae; Brucella.
OX NCBI_TaxID=29459, 29461;
RN [1]
SEQUENCE FROM N.A.
SPCIES=B. melitensis; STRAIN=16M / ATCC 23456 / Biotype 1;
EX MEDLINE=20020109; PubMed=11756688;
RA Delvecchio V.G., Kapral V., Redkar R.J., Patra G., Mujar C., Los T.,
RA Ivanova N., Anderson I., Shattacharya A., Lykidis A., Reznik G.,
RA Jablonski L., Larsen L., D'Souza M., Bernal A., Mazur M., Golsman E.,
RA Selkov E., Elzer P.H., Hagius S., O'Callaghan D., Letesson J.-J.,
RA Haselkorn R., Kyripides N., Overbeek R.;
RT "The genome sequence of the facultative intracellular pathogen
RT Brucella melitensis."
EL Proc. Natl. Acad. Sci. U.S.A. 99:443-448(2002).
RN [2]
SEQUENCE FROM N.A.
RC SPCIES=B. suis; STRAIN=1330 / Biovar 1;
RX MEDLINE=22247741; PubMed=12271122;
RA Paulsen I.T., Seshadri R., Nelson K.E., Eisen J.A., Heidelberg J.F.,
RA Read T.D., Dodson R.J., Nayana L., Brinkac L.M., Beanan M.J.,
RA Daugherty S.C., Deboy R.T., Durkin A.S., Kolonay J.F., Madupu R.,
RA Nelson W.C., Ayodeji B., Krahl M., Shetty J., Malek J., Van Aken S.E.,
RA Riedmuller S., Letellin H., Gill S.R., White O., Salzberg S.L.,
RA Hoover D.L., Lindler L.E., Halling S.M., Boyle S.M., Fraser C.M.;
RT "The Brucella suis genome reveals fundamental similarities between
RT animal and plant pathogens and symbionts."
RL Proc. Natl. Acad. Sci. U.S.A. 99:13148-13153(2002).
CC -!- FUNCTION: Plays an important role in the de novo pathway of purine
CC -!- nucleotide biosynthesis.
CC -!- CATALYTIC ACTIVITY: GTP + IMP + L-aspartate = GDP + phosphate +
CC adenylosuccinate.
CC -!- COFACTOR: Binds 1 magnesium ion per subunit (By similarity).
CC -!- PATHWAY: AMP biosynthesis; first committed step.
CC -!- SUBUNIT: Homodimer (By similarity).
CC -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -!- SIMILARITY: Belongs to the adenylosuccinate synthetase family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AE009477; AAL51532.1; ALT_INIT.
```

```
DR EMBL; AE014461; AAN30583.1; -.
DR PIR; A13295; A13295.
DR TIGR; BR1683; -.
DR HAMAP; MF 00011; -.
DR InterPro; IPR001114; Asucc synthase.
DR Pfam; PF00709; Adenylosucc synt; 1.
DR ProDom; PD001188; Asucc synthase; 3.
DR TIGRFAMs; TIGR00184; PUZA; 1.
DR PROSITE; PS01266; ADENYLOSUCIN SYN 1; 1.
DR PROSITE; PS00513; ADENYLOSUCIN SYN 2; 1.
KW purine biosynthesis; Ligase; GTP-binding; Metal-binding; Magnesium;
KW Complete proteome.
FT NP_BIND 103 109 GTP (POTENTIAL).
FT ACT_SITE 231 231 BY SIMILARITY.
FT ACT_SITE 238 238 BY SIMILARITY.
FT METAL 13 13 MAGNESIUM (BY SIMILARITY).
FT METAL 40 40 MAGNESIUM (VIA CARBONYL OXYGEN) (BY
FT SIMILARITY).
SQ SEQUENCE 429 AA; 46561 MW; DF90E83E30690FF2 CRC64;

Query Match 2.9%; Score 7; DB 1; Length 429;
Best Local Similarity 100.0%; Pred. No. 36;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 149 GYVLGIT 155
    |||||
Db 259 GYVLGIT 265

RESULT 15
PURA RHIL0
ID PURA RHIL0 STANDARD; PRT; 432 AA.
AC Q98F97;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Adenylosuccinate synthetase (EC 6.3.4.4) (IMP--aspartate ligase)
DE (AdSS) (AMPSase).
GN PURA OR Mlu3873.
OS Rhizobium loti (Mesorhizobium loti).
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC Phyllobacteriaceae; Mesorhizobium.
OX NCBI_TaxID=381;
RN [1]
SEQUENCE FROM N.A.
RC STRAIN=NAFF303099; PubMed=11214968;
RX Kaneko T., Nakamura Y., Sato S., Asamizu E., Kato T., Sasamoto S.,
RA Watanabe A., Ideawara K., Ishikawa A., Kawashima K., Kimura T.,
RA Kishida Y., Kiyokawa C., Kohara M., Matsumoto M., Matsuno A.,
RA Mochizuki Y., Nakayama S., Nakazaki N., Shimo S., Sugimoto M.,
RA Takeuchi C., Yamada M., Tabata S.;
RT "Complete genome structure of the nitrogen-fixing symbiotic bacterium
RT Mesorhizobium loti."
RL DNA Res. 7:331-338(2000).
CC -!- FUNCTION: Plays an important role in the de novo pathway of purine
CC -!- nucleotide biosynthesis.
CC -!- CATALYTIC ACTIVITY: GTP + IMP + L-aspartate = GDP + phosphate +
CC adenylosuccinate.
CC -!- COFACTOR: Binds 1 magnesium ion per subunit (By similarity).
CC -!- PATHWAY: AMP biosynthesis; first committed step.
CC -!- SUBUNIT: Homodimer (By similarity).
CC -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -!- SIMILARITY: Belongs to the adenylosuccinate synthetase family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AE014461; AAN30583.1; -.
CC PIR; A13295; A13295.
CC TIGR; BR1683; -.
CC HAMAP; MF 00011; -.
CC InterPro; IPR001114; Asucc synthase.
CC Pfam; PF00709; Adenylosucc synt; 1.
CC ProDom; PD001188; Asucc synthase; 3.
CC TIGRFAMs; TIGR00184; PUZA; 1.
CC PROSITE; PS01266; ADENYLOSUCIN SYN 1; 1.
CC PROSITE; PS00513; ADENYLOSUCIN SYN 2; 1.
CC KW purine biosynthesis; Ligase; GTP-binding; Metal-binding; Magnesium;
CC KW Complete proteome.
CC FT NP_BIND 103 109 GTP (POTENTIAL).
CC FT ACT_SITE 231 231 BY SIMILARITY.
CC FT ACT_SITE 238 238 BY SIMILARITY.
CC FT METAL 13 13 MAGNESIUM (BY SIMILARITY).
CC FT METAL 40 40 MAGNESIUM (VIA CARBONYL OXYGEN) (BY
CC SIMILARITY).
CC SQ SEQUENCE 429 AA; 46561 MW; DF90E83E30690FF2 CRC64;

Query Match 2.9%; Score 7; DB 1; Length 429;
Best Local Similarity 100.0%; Pred. No. 36;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 149 GYVLGIT 155
    |||||
Db 259 GYVLGIT 265
```

```

DR EMBL; AP003002; BAB50670.1; -.
DR HAMAP; MF 00011; -.
DR InterPro; IPR001114; Asucc_synthetase.
DR Pfam; PF00709; Adenylsucc_synth; 1.
DR ProDom; PD001188; Asucc_synthetase; 1.
DR TIGRFAMs; TIGR00184; purA; 1.
DR PROSITE; PS01266; ADENYLOSUCIN SYN 1; 1.
DR PROSITE; PS00513; ADENYLOSUCIN SYN 2; 1.
DR Purine biosynthesis; Ligase; GTP-binding; Metal-binding; Magnesium;
KW Complete proteome.
FT NP_BIND 12 18 GTP (POTENTIAL).
FT ACT_SITE 143 143 BY SIMILARITY.
FT ACT_SITE 150 150 BY SIMILARITY.
FT METAL 13 13 MAGNESIUM (BY SIMILARITY).
FT METAL 40 40 MAGNESIUM (VIA CARBONYL OXYGEN) (BY
SIMILARITY).
SQ SEQUENCE 432 AA; 46671 MW; 8726979D4C5E5D07 CRC64;

Query Match 2.9%; Score 7; DB 1; Length 432;
Best Local Similarity 100.0%; Pred. No. 37;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 149 GYVLGIT 155
DB 262 GYVLGIT 268

RESULT 16
SYE2 COXBU STANDARD; PRT; 464 AA.
AC Q83B76;
DT 15-MAR-2004 (Rel. 43, Created)
DT 15-MAR-2004 (Rel. 43, Last sequence update)
DE Glutamy1-tRNA synthetase 2 (EC 6.1.1.17) (Glutamate--tRNA ligase 2)
DE (GluRS 2)
GN GLTX2 OR GLTX-2 OR CBUL488.
OS Coccidia burnetii
OC Bacteria; Proteobacteria; Gammaproteobacteria; Legionellales;
OC Coxiellaceae; Coxiella.
OX NCBI_TaxID=777;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Nine Mile phase I / RSA 493;
RX MEDLINE=22608657; PubMed=12704232;
RA Seshadri R., Paulsen I.T., Eisen J.A., Read T.D., Nelson K.E.,
Nelson W.C., Ward N.L., Tettelin H., Daviden T.M., Beanan M.J.,
DeBoy R.T., Daugherty S.C., Brinkac L.M., Madupu R., Dodson R.J.,
Khoury H.M., Lee K.H., Carty H.A., Scanlan D., Heinzen R.A.,
Thompson H.A., Samuel J.E., Fraser C.M., Heidelberg J.F.;
RT "Complete genome sequence of the Q-fever pathogen, Coxiella
burnetii.";
RL Proc Natl. Acad. Sci. U.S.A. 100:5455-5460(2003).
CC -!- CATALYTIC ACTIVITY: ATP + L-glutamate + tRNA(Glu) = AMP +
diphosphate + L-glutamyl-tRNA(Glu).
CC -!- SUBUNIT: Monomer (By similarity).
CC -!- SUBCELLULAR LOCATION: Cytoplasmic.
CC -!- SIMILARITY: Belongs to class-I aminocyl-tRNA synthetase family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC
CC EMBL; AE016964; AAC90985.1; -.
DR TIGR; CBUL488; -.
DR HAMAP; MF 00022; -.
DR InterPro; IPR004527; Gltx_bact.
DR InterPro; IPR000924; Glu_tRNA-synt_1c.
DR InterPro; IPR008925; tRNA-synt_bind.

```

```

DR InterPro; IPR001412; tRNA-synt_1.
DR Pfam; PF00749; tRNA-synt_1c; 1.
DR PRINTS; PRO0987; TRNASYNTHGLU.
DR TIGRFAMs; TIGR00464; Gltx_bact; 1.
DR PROSITE; PS00178; AA_TRNA_LIGASE_1; 1.
KW Aminoacyl-tRNA synthetase; Protein biosynthesis; Ligase; ATP-binding;
KW Complete proteome.
FT SITE 7 17 "HIGH" REGION.
FT SITE 248 252 "KMSKS" REGION.
FT BINDING 251 251 ATP (BY SIMILARITY).
SQ SEQUENCE 464 AA; 52631 MW; 9DD5BD42123B4F6B CRC64;

Query Match 2.9%; Score 7; DB 1; Length 464;
Best Local Similarity 100.0%; Pred. No. 39;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 100 EGPGADE 106
DB 68 EGPGADE 74

RESULT 17
FTSA_PORGI STANDARD; PRT; 479 AA.
ID FTSA_PORGI
AC O07827;
DT 15-JUL-1998 (Rel. 36, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Cell division protein ftsA.
GN FTSA OR PG0583.
OS Porphyromonas gingivalis (Bacteroides gingivalis).
OC Bacteria; Bacteroidetes; Bacteroides (Class); Bacteroidales;
OC Porphyromonadaceae; Porphyromonas.
OX NCBI_TaxID=837;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=381;
RA Akifusa S., Tamura H., Anusai T., Takehara T.;
RL Submitted (JUN-1997) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=W83;
RX MEDLINE=22829867; PubMed=12949112;
RA Nelson K.E., Fleischmann R.D., DeBoy R.T., Paulsen I.T., Fouts D.E.,
Eisen J.A., Daugherty S.C., Dodson R.J., Durkin A.S., Gwinn M.,
Haft D.H., Kolonay J.F., Nelson W.C., Mason T., Tallon L., Gray J.,
Granger D., Tettelin H., Dong H., Galvin J.L., Duncan M.J.,
Dewhirst F.E., Fraser C.M.;
RT "Complete genome sequence of the oral pathogenic bacterium
Porphyromonas gingivalis strain W83.";
RL J. Bacteriol. 185:5591-5601(2003).
CC -!- FUNCTION: This protein may be involved in anomalous filament
growth (By similarity). May be a component of the septum. It
may interact with ftsZ (By similarity).
CC -!- SIMILARITY: Belongs to the ftsA/mreB family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC
CC EMBL; AB004555; BAA20534.1; -.
DR EMBL; AE017174; AAC65770.1; -.
DR TIGR; PG0583; -.
DR InterPro; IPR003494; ftsA.
DR Pfam; PF02491; ftsA; 2.
DR TIGRFAMs; TIGR01174; ftsA; 1.
KW Cell division; Cell shape; Complete proteome.
FT CONFLICT 185 185 T -> A (IN REF. 1).
FT CONFLICT 369 369 E -> G (IN REF. 1).

```



```
FT CONFLICT 405 405 I -> T (IN REF. 1).
FT CONFLICT 418 418 G -> D (IN REF. 1).
FT CONFLICT 421 421 T -> A (IN REF. 1).
FT CONFLICT 441 441 K -> R (IN REF. 1).
FT CONFLICT 444 444 V -> L (IN REF. 1).
FT CONFLICT 457 479 MISSING (IN REF. 1).
SQ SEQUENCE 479 AA; 53909 MW; 199AAA73ED8561C9 CRC64;

Query Match 2.9%; Score 7; DB 1; Length 479;
Best Local Similarity 100.0%; Pred. No. 40;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 163 IGAGIIL 169
Db 321 IGAGIIL 327

RESULT 18
EMBL-ECOLI STANDARD; PRT; 512 AA.
AC P27304; F77725;
DT 01-AUG-1992 (Rel. 23, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Multidrug resistance protein B
GN EMR B Or B2686 OR Z3987 OR ECS3548.
OS Escherichia coli, and
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=562, 83334;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=93028382; PubMed=1403590;
RA Lomovekaya O., Lewis K.;
RT "Emr, an Escherichia coli locus for multidrug resistance.";
RL P roc. Natl. Acad. Sci. U.S.A. 89:8938-8942(1992).
[2]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / MG1655;
RX MEDLINE=97426617; PubMed=9278503;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12.";
RL Science 277:1453-1474(1997).
[3]
RP SEQUENCE FROM N.A.
RC STRAIN=K12;
RX MEDLINE=97349980; PubMed=9205837;
RA Yamamoto Y., Aiba H., Baba T., Hayashi K., Inada T., Isono K.,
RA Itoh T., Kimura S., Kitagawa M., Makino K., Miki T., Mitsushashi N.,
RA Mizobuchi K., Mori H., Nakade S., Nakamura Y., Nishimoto H.,
RA Oshima T., Oyama S., Saito N., Sampei G., Satoh Y., Sivasubram S.,
RA Tagami H., Takahashi H., Takeda J., Takemoto K., Uehara K., Wada C.,
RA Yanagata S., Horiuchi T.;
RT "Construction of a contiguous 874-kb sequence of the Escherichia coli
RT - K12 genome corresponding to 50.0-68.8 min on the linkage map and
RT analysis of its sequence features.";
RL DNA Res. 4:91-113(1997).
[4]
RP SEQUENCE FROM N.A.
RC STRAIN=O157:H7 / EDL933 / ATCC 700927;
RX MEDLINE=21074935; PubMed=11206551;
RA Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,
RA Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,
RA Posfai G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,
RA Grotbeck B.J., Davis N.W., Lim A., Dimalanta E.T., Potamouais K.,
RA Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,
RA Welch R.A., Blattner F.R.;
RT "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7.";
RL Nature 409:529-533(2001).
```

[5]  
SEQUENCE FROM N.A.  
STRAIN=O157:H7 / RMD 050952;  
MEDLINE=21156231; PubMed=11358796;  
RX Hayashi T., Makino K., Onishi M., Kurokawa K., Ishii K., Yokoyama K.,  
RA Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,  
RA Iida T., Takami H., Honda T., Sasakawa C., Ogasawara N., Yasunaga T.,  
RA Kihara S., Shiba T., Hattori M., Shinagawa H.;  
RT "Complete genome sequence of enterohaemorrhagic Escherichia coli  
RT O157:H7 and genomic comparison with a laboratory strain K-12.";  
RL DNA Res. 8:11-22(2001).  
CC -1- FUNCTION: TRANSLOCASE THAT CONFERS RESISTANCE TO SUBSTANCES OF  
CC HIGH HYDROPHOBICITY.  
CC -1- SUBCELLULAR LOCATION: Integral membrane protein. Inner membrane.  
CC -1- SIMILARITY: BELONGS TO THE MAJOR FACILITATOR FAMILY (ALSO KNOWN  
CC AS THE DRUG RESISTANCE TRANSLOCASE FAMILY).  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; M86657; AAA23725.1; --  
CC EMBL; AE000353; AAC75733.1; --  
CC EMBL; D90891; BAA16548.1; --  
CC EMBL; D90892; BAA16553.1; --  
CC EMBL; AE005497; AAG57795.1; --  
CC EMBL; AP002562; BAB36971.1; --  
CC PIR; D91072; D91072.  
CC PIR; G65048; G65048.  
CC EcoGene; EG11439; emrB  
CC InterPro; IPR004638; Efflux\_EmrB.  
CC InterPro; IPR007114; MFS.  
CC InterPro; IPR005828; Sub\_transporter.  
CC Pfam; PF0083; sugar\_tr; 1.  
CC TIGRFAMs; TIGR00711; efflux\_EmrB; 1.  
CC PROSITE; PS50850; MFS; 1.  
CC Antibiotic resistance; Transport; Transmembrane; Inner membrane;  
CC Complete proteome.  
KW DOMAIN 1 12 CYTOPLASMIC (POTENTIAL).  
FT TRANSMEM 13 38 POTENTIAL.  
FT DOMAIN 39 51 EXTRACELLULAR (POTENTIAL).  
FT TRANSMEM 52 72 POTENTIAL.  
FT DOMAIN 73 81 CYTOPLASMIC (POTENTIAL).  
FT TRANSMEM 82 100 POTENTIAL.  
FT DOMAIN 101 109 EXTRACELLULAR (POTENTIAL).  
FT TRANSMEM 110 128 POTENTIAL.  
FT DOMAIN 129 136 CYTOPLASMIC (POTENTIAL).  
FT TRANSMEM 137 159 POTENTIAL.  
FT DOMAIN 160 164 EXTRACELLULAR (POTENTIAL).  
FT TRANSMEM 165 189 POTENTIAL.  
FT DOMAIN 190 202 CYTOPLASMIC (POTENTIAL).  
FT TRANSMEM 203 223 POTENTIAL.  
FT DOMAIN 224 233 EXTRACELLULAR (POTENTIAL).  
FT TRANSMEM 234 249 POTENTIAL.  
FT DOMAIN 250 271 CYTOPLASMIC (POTENTIAL).  
FT TRANSMEM 272 295 POTENTIAL.  
FT DOMAIN 296 305 EXTRACELLULAR (POTENTIAL).  
FT TRANSMEM 306 329 POTENTIAL.  
FT DOMAIN 330 335 CYTOPLASMIC (POTENTIAL).  
FT TRANSMEM 336 355 POTENTIAL.  
FT DOMAIN 356 363 EXTRACELLULAR (POTENTIAL).  
FT TRANSMEM 364 387 POTENTIAL.  
FT DOMAIN 388 407 CYTOPLASMIC (POTENTIAL).  
FT TRANSMEM 408 428 POTENTIAL.  
FT DOMAIN 429 481 EXTRACELLULAR (POTENTIAL).  
FT TRANSMEM 482 504 POTENTIAL.  
FT DOMAIN 505 512 CYTOPLASMIC (POTENTIAL).  
FT TRANSMEM 510 512 POLY-GLY.  
FT DOMAIN 512 512  
FT CONFLICT 325 325 G -> A (IN REF. 1).

FT CONFLICT 501 501 P -> A (IN REF. 1).  
SQ SEQUENCE 512 AA; 55624 MW; 070334568B303EFF CRC64;

Query Match 2.9%; Score 7; DB 1; Length 512;  
Best Local Similarity 100.0%; Pred. No. 42;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 34 GLASAPV 40  
| | | | |  
DB 307 GLASAPV 313

RESULT 19  
TY3H HUMAN  
ID TY3H HUMAN STANDARD; PRT; 528 AA.  
AC P07101; Q15585; Q15588; Q15589;  
DT 01-APR-1988 (Rel. 07, Created)  
DT 01-JUL-1989 (Rel. 11, Last sequence update)  
DT 10-OCT-2003 (Rel. 42, Last annotation update)  
DE Tyrosine 3-monoxygenase (EC 1.14.16.2) (Tyrosine 3-hydroxylase) (TH).  
DS TH OR TYH.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
[1]  
RP SEQUENCE FROM N.A. (ISOFORMS 1; 2; 3 AND 4).  
RX MEDLINE=87173054; PubMed=2892428;  
RA Grima B., Lamouroux A., Bonl C., Julien J.-P., Javoy-Agid F.,  
RA Maillet J.;  
RT "A single human gene encoding multiple tyrosine hydroxylases with  
RT different predicted functional characteristics";  
RL Nature 326:707-711(1987).  
RN [2]  
RP SEQUENCE FROM N.A. (ISOFORM 3).  
RX MEDLINE=87316931; PubMed=2888085;  
RA Kobayashi K., Kaneda N., Ichinose H., Kishi F., Nakazawa A.,  
RA Kurosawa Y., Fujita K., Nagatsu T.;  
RT "Isolation of a full-length cDNA clone encoding human tyrosine  
RT hydroxylase type 3";  
RL Nucleic Acids Res. 15:6733-6733(1987).  
RN [3]  
RP SEQUENCE OF 1-94 FROM N.A. (ISOFORM 3).  
RX MEDLINE=87298614; PubMed=2887165;  
RA Kaneda N., Kobayashi K., Ichinose H., Kishi F., Nakazawa A.,  
RA Kurosawa Y., Fujita K., Nagatsu T.;  
RT "Isolation of a novel cDNA clone for human tyrosine hydroxylase:  
RT alternative RNA splicing produces four kinds of mRNA from a single  
RT gene";  
RL Biochem. Biophys. Res. Commun. 146:971-975(1987).  
RN [4]  
RP SEQUENCE FROM N.A. (ISOFORMS 1; 2; 3 AND 4).  
RX MEDLINE=89008200; PubMed=2902075;  
RA Kobayashi K., Kaneda N., Ichinose H., Kishi F., Nakazawa A.,  
RA Kurosawa Y., Fujita K., Nagatsu T.;  
RT "Structure of the human tyrosine hydroxylase gene: alternative  
RT splicing from a single gene accounts for generation of four mRNA  
RT types";  
RL J. Biochem. 103:907-912(1988).  
RN [5]  
RP PARTIAL SEQUENCE FROM N.A.  
RX MEDLINE=88117543; PubMed=2892893;  
RA le Bourdelles B., Boularand S., Boni C., Horellou P., Dumas S.,  
RA Grima B., Maillet J.;  
RT "Analysis of the 5' region of the human tyrosine hydroxylase gene:  
RT combinatorial patterns of exon splicing generate multiple regulated  
RT tyrosine hydroxylase isoforms";  
RL J. Neurochem. 50:988-991(1988).  
RN [6]  
RP SEQUENCE OF 1-30 FROM N.A.  
RX MEDLINE=88213428; PubMed=2896667;  
RA Gims E.I., Rehavi M., Martin B.M., Weller M., O'Malley K.L.,  
RA Lamarca M.E., McAllister C.G., Paul S.W.;

RT "Expression of human tyrosine hydroxylase cDNA in invertebrate cells  
RT using a baculovirus vector";  
RL J. Biol. Chem. 263:7406-7410(1988).  
RN [7]  
RP VARIANT SEGAWA SYNDROME LYS-412.  
RX MEDLINE=95113436; PubMed=7814018;  
RA Luedecke B., Dworniczak B., Bartholome K.;  
RT "A point mutation in the tyrosine hydroxylase gene associated with  
RT Segawa's syndrome";  
RL Hum. Genet. 95:123-125(1995).  
RN [8]  
RP VARIANT MET-112.  
RX MEDLINE=95305895; PubMed=7789962;  
RA Luedecke B., Bartholome K.;  
RT "Frequent sequence variant in the human tyrosine hydroxylase gene";  
RL Hum. Genet. 95:716-716(1995).  
RN [9]  
RP CHARACTERIZATION OF VARIANT SEGAWA SYNDROME LYS-412.  
RX MEDLINE=96133297; PubMed=8528210;  
RA Knappskog P.M., Flatmark T., Maillet J., Luedecke B., Bartholome K.;  
RT "Recessively inherited L-DOPA-responsive dystonia caused by a point  
RT mutation (Q381K) in the tyrosine hydroxylase gene";  
RL Hum. Mol. Genet. 4:1209-1212(1995).  
RN [10]  
RP CHARACTERIZATION OF VARIANT SEGAWA SYNDROME PRO-236.  
RX MEDLINE=96414313; PubMed=8817341;  
RA Luedecke B., Knappskog P.M., Clayton P.T., Surtees R.A.H.,  
RA Clelland J.D., Heales S.J.R., Brand M.P., Bartholome K., Flatmark T.;  
RT "Recessively inherited L-DOPA-responsive parkinsonism in infancy  
RT caused by a point mutation (L203P) in the tyrosine hydroxylase gene";  
RL Hum. Mol. Genet. 5:1023-1028(1996).  
RN [11]  
RP VARIANT SEGAWA SYNDROME PRO-236, AND VARIANT MET-112.  
RX MEDLINE=98273932; PubMed=9613851;  
RA Kunugi H., Kawada Y., Hattori M., Ueki A., Otsuka M., Nanko S.;  
RT "Association study of structural mutations of the tyrosine hydroxylase  
RT gene with schizophrenia and Parkinson's disease";  
RL Am. J. Med. Genet. 81:131-133(1998).  
RN [12]  
RP VARIANT MET-499.  
RX MEDLINE=98425595; PubMed=9754624;  
RA Ishiguro H., Arinami T., Saito T., Akazawa S., Enomoto M.,  
RA Mitushio H., Fujishiro H., Tada K., Akimoto Y., Mifune H.,  
RA Shozuka S., Hamaguchi H., Toru M., Shibuya H.;  
RT "Systematic search for variations in the tyrosine hydroxylase gene and  
RT their associations with schizophrenia, affective disorders, and  
RT alcoholism";  
RL Am. J. Med. Genet. 81:388-396(1998).  
RN [13]  
RP VARIANT SEGAWA SYNDROME HIS-233.  
RX MEDLINE=98365985; PubMed=9703425;  
RA van den Heuvel L.P.W.J., Luiten B., Smeitink J.A.M.,  
RA de Rijk-van Andel J.F., Hyland K., Steenbergen-Spanjers G.C.H.,  
RA Janssen R.J.T., Wevers R.A.;  
RT "A common point mutation in the tyrosine hydroxylase gene in autosomal  
RT recessive L-DOPA-responsive dystonia in the Dutch population";  
RL Hum. Genet. 102:644-646(1998).  
RN [14]  
RP VARIANT MET-112.  
RX MEDLINE=99318093; PubMed=10391209;  
RA Cargill M., Altshuler D., Ireland J., Sklar P., Ardlie K., Patil N.,  
RA Shaw N., Lane C.R., Lim E.P., Kalyanaram N., Nemesh J., Ziaugra L.,  
RA Friedland L., Rolfe A., Warrington J., Lipshutz R., Daley G.Q.,  
RA Lander E.S.;  
RT "Characterization of single-nucleotide polymorphisms in coding regions  
RT of human genes";  
RL Nat. Genet. 22:231-238(1999).  
RN [15]  
RP ERRATUM.  
RA Cargill M., Altshuler D., Ireland J., Sklar P., Ardlie K., Patil N.,  
RA Shaw N., Lane C.R., Lim E.P., Kalyanaram N., Nemesh J., Ziaugra L.,  
RA Friedland L., Rolfe A., Warrington J., Lipshutz R., Daley G.Q.,  
RA Lander E.S.;

RL Nat. Genet. 23:373-373(1999).  
RN [16]  
RP VARIANTS SEGAWA SYNDROME PRO-276; MET-314; HIS-337 AND MET-494.  
RX MEDLINE=21141790; PubMed=11246459;  
RA Swans R.J.M., Rondot P., Renier W.O., Van Den Heuvel L.P.W.J.,  
RA Steenberg-Spanjers G.C.H., Wevers R.A.,  
RT "Four novel mutations in the tyrosine hydroxylase gene in patients  
RT with infantile parkinsonism".  
RL Ann. Hum. Genet. 64:25-31(2000).  
CC -!- FUNCTION: Plays an important role in the physiology of adrenergic  
CC neurones.  
CC -!- CATALYTIC ACTIVITY: L-tyrosine + tetrahydropteridine + O(2) = 3,4-  
CC dihydroxy-L-phenylalanine + dihydropteridine + H(2)O.  
CC -!- COFACTOR: Ferrous ion.  
CC -!- ENZYME REGULATION: Phosphorylation leads to an increase in the  
CC catalytic activity.  
CC -!- PATHWAY: Catecholamine biosynthesis; first step.  
CC -!- ALTERNATIVE PRODUCTS:  
CC Event=Alternative splicing; Named isoforms=4;  
CC Name=3;  
CC IsoId=P07101-1; Sequence=VSP\_000542;  
CC Name=1;  
CC IsoId=P07101-2; Sequence=VSP\_000543;  
CC Name=2;  
CC IsoId=P07101-3; Sequence=VSP\_000544;  
CC Name=4;  
CC IsoId=P07101-4; Sequence=VSP\_000541; VSP\_000542;  
CC -!- TISSUE SPECIFICITY: Mainly expressed in the brain and adrenal  
CC glands.  
CC -!- DISPAR: Defects in TH are the cause of autosomal recessive Segawa  
CC syndrome [MIM:605407]; also known as DOPA-responsive dystonia.  
CC Typically, it begins in childhood or adolescence with progressive  
CC difficulty in walking and in some cases, spasticity. Some cases  
CC present with parkinsonian symptoms in infancy and are referred to  
CC as autosomal recessive infantile parkinsonism.  
CC -!- SIMILARITY: Belongs to the biotin-dependent aromatic amino acid  
CC hydroxylase family.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL: Y00414; CAA68472.1; ALT\_SEQ.  
CC EMBL: M17589; AAA61179.1; -.  
CC EMBL: X05290; CAA28908.1; -.  
CC EMBL: M24791; AAA61172.1; -.  
CC EMBL: M24787; AAA61173.1; JOINED.  
CC EMBL: M24789; AAA61173.1; JOINED.  
CC EMBL: M24791; AAA61170.1; -.  
CC EMBL: M24787; AAA61170.1; JOINED.  
CC EMBL: M20911; AAA61167.1; -.  
CC PIR: A30002; WHUY4.  
CC HSSP: P04177; LTOH.  
CC Genew: HGNC:11782; TH.  
CC XTM: 191290; -.  
CC NIM: 605407; -.  
CC GO: GO:0004511; P-tyrosine 3-monooxygenase activity; TAS.  
CC GO: GO:0006584; P-catecholamine metabolism; TAS.  
CC GO: GO:0007345; P-embryogenesis and morphogenesis; TAS.  
CC InterPro: IPR001273; Aaa\_hydroxylase.  
CC Pfam: PF00351; biotin\_H\_1.  
CC PRINTS: PR00372; FWHYDXLASE.  
CC ProDom: PD002559; Aaa\_hydroxylase; 1.  
Query Match 2.9%; Score 7; DB 1; Length 528;  
Best Local Similarity 100.0%; Pred.No. 43;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 121 EAAAVQP 127  
Db 446 EAAAVQP 452  
RESULT 20  
URE1\_PROMI STANDARD; PRT; 567 AA.  
ID URE1\_PROMI  
AC P17086;  
DT 01-AUG-1990 (Rel. 15, Created)  
DT 01-AUG-1990 (Rel. 15, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE Urease alpha subunit (EC 3.5.1.5) (Urea amidohydrolase).  
GN UREC.  
OS Proteus mirabilis.  
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
OC Enterobacteriaceae; Proteus.  
OX NCBI\_TaxID=584;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX STRAIN=HI4320;  
RC MEDLINE=90078080; PubMed=2687233;  
RA Jones B.D., Mobley H.L.T.;  
RT "Proteus mirabilis urease: nucleotide sequence determination and  
RT comparison with jack bean urease".  
RL J. Bacteriol. 171:5414-5422(1989).  
CC -!- CATALYTIC ACTIVITY: Urea + H(2)O = CO(2) + 2 NH(3).  
CC -!- COFACTOR: Binds 2 nickel ions per subunit (By similarity).  
CC -!- SUBUNIT: (Alpha, beta, gamma) (3) (By similarity).  
CC -!- PFM: Lys-217 is carbamylated. The carbamoyl group provides the  
CC ligands for the two nickel ions (By similarity).  
CC -!- SIMILARITY: Belongs to the urease family.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL: M31934; AAA25669.1; -.  
CC PIR: D43719; D43719.  
CC HSSP: P18314; 1FWE.  
CC MEROPS: M38.UNW; -.  
CC InterPro: IPR006680; Amidohydro 1.  
CC InterPro: IPR005848; Pept M38 urease.  
CC Pfam: PF01973; Urease\_alpha.  
CC Pfam: PF00449; Urease; 1; 1.  
CC PIRSF: PIRSF001226; Urease\_alpha; 1.  
CC PRINTS: PR01752; UREASE.  
CC PROSITE: PS00145; UREASE\_2; 1.  
CC PROSITE: PS01120; UREASE\_1; 1.  
CC Hydrolase; Metal-binding; Nickel.  
CC FT METAL 134 134 NICKEL 2 (BY SIMILARITY).  
CC FT METAL 136 136 NICKEL 2 (BY SIMILARITY).  
CC FT METAL 217 217 NICKEL 1 AND 2 (BY SIMILARITY).  
CC FT METAL 246 246 NICKEL 1 (BY SIMILARITY).  
CC FT METAL 272 272 NICKEL 1 (BY SIMILARITY).  
CC FT METAL 360 360 NICKEL 2 (BY SIMILARITY).  
CC FT ACT SITE 320 320 BY SIMILARITY.  
CC SQ SEQUENCE 567 AA; 61012 MW; 67794FDS495A9E1 CRC64;  
Query Match 2.9%; Score 7; DB 1; Length 567;  
Best Local Similarity 100.0%; Pred.No. 46;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 66 EAGVPEX 72  
Db 498 EAGVPEX 504

DR TIGRFAMS; TIGR00355; purH; 1.  
 KW Purine biosynthesis; Transferase; Hydrolase; Multifunctional enzyme.  
 SQ SEQUENCE 591 AA; 65282 MW; DEC0684BFED7CA7 CRC64;  
 Query Match 2.9%; Score 7; DB 1; Length 591;  
 Best Local Similarity 100.0%; Pred No. 48;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 177 KDLXEQH 183  
 DB 87 KDLXEQH 93  
 RESULT 22  
 PU91\_YEAST STANDARD; PRT; 591 AA.  
 ID PU91\_YEAST  
 AC P54113;  
 DT 01-OCT-1996 (Rel. 34, Created)  
 DT 01-OCT-1996 (Rel. 34, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE Bifunctional purine biosynthesis protein ADE16 [Includes:  
 DE Phosphoribosylamidazolecarboxamide formyltransferase (EC 2.1.2.3)  
 DE (AICAR transferase); IMP cyclohydrolase (EC 3.5.4.10) (Inosinase)  
 DE (IMP synthetase) (ATIC)]  
 GN ADE16 OR YLR028C.  
 OS Saccharomyces cerevisiae (Baker's yeast).  
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
 OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.  
 OX NCBI\_TaxID=4932;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=S288C / AB972;  
 RX MEDLINE=97313267; PubMed=9169871;  
 RA Johnston M., Hillier L., Kiles L., Dubois E., Duesterhoef A.,  
 RA Benes V., Brueckner M., Delius H., Hebling U., Heumann K.,  
 RA Entian K.-D., Floeth M., Goffeau A., Hilger F., Kleine K., Koetter P.,  
 RA Heuss-Neitzel D., Hilbert H., Hilger F., Miesha T., Moestl D.,  
 RA Louis B.J., Messenguy F., Mewes H.-W., Miosga T., Moestl D.,  
 RA Mueller-Auer S., Nentwich U., Obermaier B., Piravandi E., Pohl T.M.,  
 RA Portetle D., Purnelle B., Rechmann S., Rieger M., Rinke M., Rose M.,  
 RA Scharte M., Scherens B., Scholler P., Schwager C., Schwarz S.,  
 RA Underwood A.P., Urrestazu L.A., Vandenbol M., Verhasselt P.,  
 RA Vierderdeels F., Voet M., Volckaert G., Voss H., Wambutt R., Wedler E.,  
 RA Wedler H., Zimmermann F.K., Zollner A., Hani J., Hobeisel J.D.;  
 RA "The nucleotide sequence of Saccharomyces cerevisiae chromosome XII.";  
 RT Nature 387:87-90(1997).  
 RL Nature 387:90-93(1997).  
 CC -1- CATALYTIC ACTIVITY: 10-formyltetrahydrofolate + 5-amino-1-(5-  
 CC phospho-D-ribosyl)imidazole-4-carboxamide = tetrahydrofolate + 5-  
 CC formamido-1-(5-phospho-D-ribosyl)imidazole-4-carboxamide.  
 CC -1- CATALYTIC ACTIVITY: IMP + H(2)O = 5-formamido-1-(5-phospho-D-  
 CC ribosyl)imidazole-4-carboxamide.  
 CC -1- PATHWAY: De novo purine biosynthesis; ninth step.  
 CC -1- PATHWAY: De novo purine biosynthesis; tenth step.  
 CC -1- SUBUNIT: Homodimer (Possible).  
 CC -1- DOMAIN: The IMP cyclohydrolase activity resides in the N-terminal  
 CC region (By similarity).  
 CC -1- SIMILARITY: Belongs to the purH family.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/  
 CC or send an email to license@sib-sib.ch).  
 CC -----  
 CC EMBL; U62402; AAB57774.1; -;  
 CC EMBL; Z73200; CAA97552.1; -;  
 CC PIR; S77707; S77707.  
 CC Germonline; 142090; -;  
 CC SGD; S0004018; ADE16.  
 CC GO; GO:0005829; C:cytosol; IDA.  
 CC GO; GO:0003937; F:IMP phosphoribosylamidazolecarboxamide for. .; IDA.  
 CC GO; GO:0004643; F:IMP phosphoribosylamidazolecarboxamide for. .; IDA.  
 CC GO; GO:0006189; P:de novo IMP biosynthesis; IDA.  
 CC GO; GO:0009060; P:aerobic respiration; IEP.  
 CC GO; GO:0007151; P:sporulation (sensu Saccharomycetes); IEP.  
 CC InterPro; IPR002695; AICARFT\_IMPCHAS.  
 CC InterPro; IPR004362; MGS like.  
 CC Pfam; PF01808; AICARFT\_IMPCHAS; 1.  
 CC Pfam; PF02142; MGS; 1.  
 CC ProDom; PD004666; AICARFT\_IMPCHAS; 1.

DR TIGRFAMS; TIGR00355; purH; 1.  
 KW Purine biosynthesis; Transferase; Hydrolase; Multifunctional enzyme.  
 SQ SEQUENCE 591 AA; 65282 MW; DEC0684BFED7CA7 CRC64;  
 Query Match 2.9%; Score 7; DB 1; Length 591;  
 Best Local Similarity 100.0%; Pred No. 48;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 177 KDLXEQH 183  
 DB 87 KDLXEQH 93  
 RESULT 22  
 PU92\_YEAST STANDARD; PRT; 592 AA.  
 ID PU92\_YEAST  
 AC P38009;  
 DT 01-OCT-1994 (Rel. 30, Created)  
 DT 01-OCT-1996 (Rel. 34, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE Bifunctional purine biosynthesis protein ADE17 [Includes:  
 DE Phosphoribosylamidazolecarboxamide formyltransferase (EC 2.1.2.3)  
 DE (AICAR transferase); IMP cyclohydrolase (EC 3.5.4.10) (Inosinase)  
 DE (IMP synthetase) (ATIC)]  
 GN ADE17 OR YMR120C OR YMR564.02C.  
 OS Saccharomyces cerevisiae (Baker's yeast).  
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
 OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.  
 OX NCBI\_TaxID=4932;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=S288C / AB972;  
 RX MEDLINE=97313268; PubMed=9169872;  
 RA Bowman S., Churcher C.M., Badcock K., Brown D., Chillingworth T.,  
 RA Connor R., Dedman K., Devlin K., Gentles S., Hamlin N., Hunt S.,  
 RA Jagels K., Lye G., Moule S., Odell C., Pearson D., Rajandream M.A.,  
 RA Rice P., Skelton J., Walsh S., Whitehead S., Barrell B.G.;  
 RA "The nucleotide sequence of Saccharomyces cerevisiae chromosome  
 RT XIII.";  
 RL Nature 387:90-93(1997).  
 CC -1- CATALYTIC ACTIVITY: 10-formyltetrahydrofolate + 5-amino-1-(5-  
 CC phospho-D-ribosyl)imidazole-4-carboxamide = tetrahydrofolate + 5-  
 CC formamido-1-(5-phospho-D-ribosyl)imidazole-4-carboxamide.  
 CC -1- CATALYTIC ACTIVITY: IMP + H(2)O = 5-formamido-1-(5-phospho-D-  
 CC ribosyl)imidazole-4-carboxamide.  
 CC -1- PATHWAY: De novo purine biosynthesis; ninth step.  
 CC -1- PATHWAY: De novo purine biosynthesis; tenth step.  
 CC -1- SUBUNIT: Homodimer (Possible).  
 CC -1- DOMAIN: The IMP cyclohydrolase activity resides in the N-terminal  
 CC region (By similarity).  
 CC -1- SIMILARITY: Belongs to the purH family.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/  
 CC or send an email to license@sib-sib.ch).  
 CC -----  
 CC EMBL; Z49273; CAA89269.1; -;  
 CC PIR; S54489; S54489.  
 CC Germonline; 142789; -;  
 CC SGD; S0004727; ADE17.

DR GO: GO:0005829; Cytosol; IDA.  
 DR GO: GO:0003937; F:IMP cyclohydrolase activity; IDA.  
 DR GO: GO:0004643; P:phosphoribosylaminoimidazolecarboxamide for . . . IDA.  
 DR GO: GO:0006189; P:de novo IMP biosynthesis; IDA.  
 DR InterPro: IPR002695; AICARFT\_IMPCHas.  
 DR InterPro: IPR004362; MGS like.  
 DR Pfam: PF01808; AICARFT\_IMPCHas; 1.  
 DR Pfam: PF02142; MGS; 1.  
 DR Pfam: PF004666; AICARFT\_IMPCHas; 1.  
 DR Tricore: TIGR00355; purH; 1.  
 KW Purine biosynthesis; Transferase; Hydrolase; Multifunctional enzyme.  
 FT CONFLICT 389 389 R -> A (IN REF. 2).  
 SQ SEQUENCE 592 AA; 65263 MW; 8ABA71761B512242 CRC64;  
 Query Match 2.9%; Score 7; DB 1; Length 592;  
 Best Local Similarity 100.0%; Pred. No. 48;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 177 KDLKEQH 183  
 DB 87 KDLKEQH 93  
 RESULT 23  
 ID L065 ARATH STANDARD; PRT; 619 AA.  
 AC Q04960; Q06737; Q42275; Q8XFF6; Q9FHC9;  
 DT 01-FEB-1995 (Rel. 31, Created)  
 DT 28-FEB-2003 (Rel. 41, Last sequence update)  
 DT 15-MAR-2004 (Rel. 43, Last annotation update)  
 DE Low-temperature-induced 65 kDa protein [Desiccation-responsive protein 29B].  
 GN LT165 OR RD29B OR AT5G52300 OR K24M7.3.  
 OS Arabidopsis thaliana (Mouse-ear cress).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;  
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.  
 ON NCBI\_TaxID:3702;  
 RX SEQUENCE FROM N.A.  
 RC STRAIN=cv. Columbia; TISSUE=Leaf;  
 RX MEDLINE=93192524; PubMed=8448363;  
 RA Nordin K., Vahala E., Palva E.T.;  
 RT "Differential expression of two related, low-temperature-induced genes in Arabidopsis thaliana (L.) Heynh.";  
 RL Plant Mol. Biol. 21:641-653(1993).  
 RN [2]  
 SQ SEQUENCE FROM N.A.  
 RC STRAIN=cv. Columbia;  
 RX MEDLINE=94143472; PubMed=8310052;  
 RA Yameguchi-Shinozaki K., Shinozaki K.;  
 RT "Arabidopsis DNA encoding two desiccation-responsive rd29 genes.";  
 RL Plant Physiol. 101:1119-1120(1993).  
 RN [3]  
 SQ SEQUENCE FROM N.A., AND INDUCTION.  
 RC STRAIN=cv. Columbia;  
 RX MEDLINE=93173109; PubMed=8437577;  
 RA Yameguchi-Shinozaki K., Shinozaki K.;  
 RT "Characterization of the expression of a desiccation-responsive rd29 gene of Arabidopsis thaliana and analysis of its promoter in transgenic plants.";  
 RL Mol. Gen. Genet. 236:331-340(1993).  
 RN [4]  
 SQ SEQUENCE FROM N.A.  
 RC STRAIN=cv. Columbia;  
 RX MEDLINE=20181125; PubMed=10718197;  
 RA Sato S., Nakamura Y., Kaneko T., Kato T., Asamizu E., Kotani H.,  
 RA Tabata S.;  
 RT "Structural analysis of Arabidopsis thaliana chromosome 5. X. Sequence features of the regions of 3,076,755 bp covered by sixty P1 and TAC clones.";  
 RL DNA Res. 7:31-63(2000).  
 RN [5]

RP SEQUENCE FROM N.A.  
 RC STRAIN=cv. Columbia;  
 RX MEDLINE=22954850; PubMed=14593172;  
 RA Yamada K., Lim J., Dale J.M., Chen H., Shinn P., Palm C.J.,  
 RA Southwick A.M., Wu H.C., Kim C.J., Nguyen M., Pham P.K., Cheuk R.F.,  
 RA Karlin-Newmann G., Liu S.X., Lam B., Sakano H., Wu T., Yu G.,  
 RA Miranda M., Quach H.L., Tripp M., Chang C.H., Lee J.M., Toriumi M.J.,  
 RA Chan M.M., Tang C.C., Onodera C.S., Deng J.M., Akiyama K., Ansari Y.,  
 RA Arakawa T., Banh J., Hango F., Bowser L., Brooks S.F., Carninci P.,  
 RA Chao Q., Choy N., Enju A., Goldsmith A.D., Gurjal M., Hansen N.F.,  
 RA Hayashizaki Y., Johnson-Hopson C., Hsuan V.W., Iida K., Karnes M.,  
 RA Khan S., Koesema E., Ishida J., Jiang P.X., Jones T., Kawai J.,  
 RA Kamiya A., Meyers C., Nakajima M., Narusaka M., Seki M., Sakurai T.,  
 RA Satou M., Tamse R., Vaysberg M., Wallender E.K., Wong C., Yamamura Y.,  
 RA Yuan S., Shinozaki K., Davis R.M., Theologis A., Ecker J.R.;  
 RT "Empirical analysis of transcriptional activity in the Arabidopsis genome.";  
 RL Science 302:842-846(2003).  
 RN [6]  
 RP SEQUENCE OF 1-121 FROM N.A.  
 RC STRAIN=cv. Columbia; TISSUE=Dry seed;  
 RA Raynal M., Grellet P., Laudie M., Meyer Y., Cooke R., Delseny M.;  
 RL Submitted (JUN-1994) to the EMBL/GenBank/DBJ databases.  
 CC -!- INDUCTION: By low temperature, and mostly by water stress or abscisic acid (ABA).  
 CC -!- SIMILARITY: BELONGS TO THE LTI165/LTI65 FAMILY.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).  
 CC -----  
 DR EMBL; X67670; CAA47902.1; -;  
 DR EMBL; D13044; BAA02375.1; -;  
 DR EMBL; AB019226; BAB10527.1; -;  
 DR EMBL; AY081282; AAL91171.1; -;  
 DR EMBL; Z34014; CAA83975.1; -;  
 DR PIR; S30153; S30153.  
 KW Multigene family; Repeat.  
 FT DOMAIN 24 27  
 FT DOMAIN 64 67  
 FT DOMAIN 404 511  
 FT REPEAT 404 408  
 FT REPEAT 442 446  
 FT REPEAT 460 464  
 FT REPEAT 490 494  
 FT REPEAT 507 511  
 FT DOMAIN 514 517  
 FT CONFLICT 24 43  
 FT CONFLICT 80 87  
 FT CONFLICT 82 83  
 FT CONFLICT 260 261  
 FT CONFLICT 414 414  
 FT CONFLICT 423 423  
 FT CONFLICT 594 619  
 FT CONFLICT 594 619  
 FT CONFLICT 619 619  
 SQ SEQUENCE 619 AA; 65971 MW; F2D2DF9C5990A00A CRC64;  
 Query Match 2.9%; Score 7; DB 1; Length 619;  
 Best Local Similarity 100.0%; Pred. No. 50;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 65 GEAGVPE 71  
 DB 261 GEAGVPE 267

```

RESULT 24
ID UAS3_HUMAN STANDARD; PRT; 661 AA.
AC P57075;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE UBASH3A protein.
GN UBASH3A
OS Homo sapiens (Human)
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORMS LONG AND SHORT).
RC TISSUE=Placenta;
RX MEDLINE=21177060; PubMed=11281453;
RA Wattenhofer M., Shibuya K., Kudoh J., Lyle R., Michaud J., Rossier C.,
RA Kawasaki K., Asakawa S., Minoshima S., Berry A., Bonne-Tamir B.,
RA Shimizu N., Antonarakis S.E., Scott H.S.;
RT "Isolation and characterization of the UBASH3A gene on 21q22.3
RT encoding a potential nuclear protein with a novel combination of
RT domains";
RL Hum. Genet. 108:140-147(2001).
CC -!- SUBCELLULAR LOCATION: Nuclear (Potential).
CC -!- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=2;
CC Name=Long;
CC IsoId=P57075-1; Sequence=Displayed;
CC Name=Short;
CC IsoId=P57075-2; Sequence=VSP_006703;
CC -!- TISSUE SPECIFICITY: Highest expression of UBASH3A in tissues
CC belonging to the immune system, including spleen, peripheral blood
CC leukocytes, and bone marrow.
CC -!- SIMILARITY: Contains 1 SH3 domain.
CC -!- SIMILARITY: Contains 1 UBA domain.
CC -!- SIMILARITY: STRONG. TO DROSOPHILA CG13604.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AJ277750; CAB91543.1; -
CC Genew; HGNC:12462; UBASH3A.
CC MIN; 605736; -
CC InterPro; IPR001452; SH3.
CC InterPro; IPR000449; UBA_domain.
CC Pfam; PF00300; PGAM; 1.
CC Pfam; PF00018; SH3; 1.
CC Pfam; PF00627; UBA; 1.
CC ProDom; PD000066; SH3; 1.
CC SMART; SM00326; SH3; 1.
CC SMART; SM00002; SH3; 1.
CC PROSITE; PS00021; SH3; 1.
CC PROSITE; PS50030; UBA; 1.
CC Nuclear protein; SH3 domain; UBA.
CC Alternative splicing.
FT DOMAIN 15 60
FT DOMAIN 276 340
FT VARSPLIC 185 223
FT GTSVSRWFPSQVPGHGNLRLNLTASFVSHVILQKY
FT -> D (in Isoform Short).
FT /FTID=VSP_006703.
SQ SEQUENCE 661 AA; 74122 MW; 60DA2E0B8CE4ABFC CRC64;

Query Match 2.9%; Score 7; DB 1; Length 661;
Best Local Similarity 100.0%; Pred. No. 52;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 128 VIGISOR 134
|||
Db 318 VIGISOR 324

```

```

RESULT 25
ID LEM3_SHEEP STANDARD; PRT; 769 AA.
AC P98109;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE P-selectin precursor (Granule membrane protein 140) (GMP-140) (PADGEM)
DE (CD62P) (Leukocyte-endothelial cell adhesion molecule 3) (LECAM3).
GN SELLP.
OS Ovis aries (Sheep)
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Caprinae; Ovis.
OX NCBI_TaxID=9940;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Heart;
RA Burns S.A., Neufeld E.J., Donady J.J.;
RL Submitted (JUL-1994) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: Ca(2+)-dependent receptor for myeloid cells that binds
CC to carbohydrates on neutrophils and monocytes. Mediates the
CC interaction of activated endothelial cells or platelets with
CC leukocytes. The ligand recognized is sialyl-Lewis X.
CC -!- SUBCELLULAR LOCATION: Type I membrane protein.
CC -!- SIMILARITY: Belongs to the selectin/LECAM family.
CC -!- SIMILARITY: Contains 1 C-type lectin family domain.
CC -!- SIMILARITY: Contains 1 EGF-like domain.
CC -!- SIMILARITY: Contains 8 Sushi (SCR) domains.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; I34270; AB59261.1; -
CC HSP; P16109; IF8B.
CC InterPro; IPR006209; EGF_like.
CC InterPro; IPR001304; Lectin C.
CC InterPro; IPR002396; Selectin.
CC InterPro; IPR000436; Sushi_SCR_CCP.
CC Pfam; PF00008; EGF; 1.
CC Pfam; PF00059; lectin_c; 1.
CC Pfam; PF00084; sushi; 8.
CC PRINTS; PR00343; SELECTIN.
CC SMART; SM00032; CCP; 8.
CC SMART; SM00034; CLECT; 1.
CC PROSITE; PS00615; C-TYPE LECTIN_1; 1.
CC PROSITE; PS50041; C-TYPE LECTIN_2; 1.
CC PROSITE; PS00022; EGF_1; 1.
CC PROSITE; PS01186; EGF_2; 1.
CC PROSITE; PS50026; EGF_3; 1.
CC Cell adhesion; Transmembrane; Glycoprotein; EGF-like domain; Lectin;
KW Selectin; Signal; Sushi; Repeat.
FT SIGNAL 1 32
FT CHAIN 33 769
FT DOMAIN 718 734
FT TRANSMEM 735 769
FT DOMAIN 735 769
FT DOMAIN 735 769
FT DOMAIN 159 195
FT DOMAIN 199 258
FT DOMAIN 261 320
FT DOMAIN 323 392
FT DOMAIN 385 444
FT DOMAIN 447 506
FT DOMAIN 509 568
FT DOMAIN 580 639

```

```
FT DOMAIN 642 701 SUSHI 8.
FT DISULFID 130 158 BY SIMILARITY.
FT DISULFID 131 150 BY SIMILARITY.
FT DISULFID 132 150 BY SIMILARITY.
FT DISULFID 163 174 BY SIMILARITY.
FT DISULFID 168 183 BY SIMILARITY.
FT DISULFID 185 194 BY SIMILARITY.
FT DISULFID 200 244 BY SIMILARITY.
FT DISULFID 230 257 BY SIMILARITY.
FT DISULFID 262 306 BY SIMILARITY.
FT DISULFID 292 319 BY SIMILARITY.
FT DISULFID 324 368 BY SIMILARITY.
FT DISULFID 354 381 BY SIMILARITY.
FT DISULFID 386 430 BY SIMILARITY.
FT DISULFID 416 443 BY SIMILARITY.
FT DISULFID 448 492 BY SIMILARITY.
FT DISULFID 478 505 BY SIMILARITY.
FT DISULFID 510 554 BY SIMILARITY.
FT DISULFID 540 567 BY SIMILARITY.
FT DISULFID 581 625 BY SIMILARITY.
FT DISULFID 611 638 BY SIMILARITY.
FT DISULFID 643 687 BY SIMILARITY.
FT DISULFID 673 700 BY SIMILARITY.
FT CARBOHYD 54 54 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 80 80 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 180 180 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 212 212 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 219 219 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 347 347 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 398 398 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 604 604 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 655 655 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 662 662 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 680 680 N-LINKED (GLCNAC. .) (POTENTIAL).
FT SITE 757 760 ENDOCYTOPOSIS SIGNAL* (PROBABLE).
FT VARIANT 566 566 S -> T.
FT VARIANT 579 579 L -> V.
SQ SEQUENCE 769 AA; 84317 MW; 23E42575D60FAB15 CRC64;

Query Match 2.9%; Score 7; DB 1; Length 769;
Best Local Similarity 100.0%; Pred. No. 60;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 19 PAPGLRC 25
Db 575 PAPGLRC 581

RESULT 26
CUL3_SCHPO
ID CUL3_SCHPO STANDARD; PRT; 785 AA.
AC Q09760; O74185;
DT 01-NOV-1995 (Rel. 32, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Cullin 3 homolog (Cul-3).
GN PCU3 OR SPAC24H6.03.
OS Schizosaccharomyces pombe (Fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC Schizosaccharomycetes.
OX NCBI_TaxID=4896;
RN SEQUENCE FROM N.A.
RP STRAIN=972;
RC MEDLINE=21848401; PubMed=11859360;
RA Wood V., Williams R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,
RA Sgouras J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,
RA Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,
RA Collins M., Connor R., Cronin A., Davis P., Feltwell T., Fraser A.,
RA Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,
RA Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagels K.,
RA James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,
RA Mooney P., Moule S., Murgall K., Murphy L., Niblett D., Odell C.,
```

```
RA Oliver K., O'Neil S., Pearson D., Quail M.A., Rabinowitsch E.,
RA Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,
RA Skelton J., Simmonds M., Squares R., Squares S., Stevens K.,
RA Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,
RA Woodward J., Volckaert G., Aert R., Robben J., Grymonprez B.,
RA Weijtens I., Vansireels E., Rieger M., Schaefer M., Mueller-Auer S.,
RA Gabel C., Fuchs M., Fritzc C., Holzer E., Moestl D., Hilbert H.,
RA Borzym K., Langer I., Beck A., Lehrach H., Wambutt R., Fumelle B.,
RA Eger P., Zimmermann W., Wedler H., Gloux S., Lelaure V., Mottier S.,
RA Goffeau A., Cadieu E., Dreano S., Gloux S., Hunt C., Moore K., Hurst S.M.,
RA Gilbert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,
RA Lucas M., Rochet M., Gaillardin C., Tallada V.A., Garzon A., Thode G.,
RA Daga R.R., Cuizado L., Jimenez J., Sanchez M., del Rey F., Benito J.,
RA Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Foraburg S.L.,
RA Ceirutti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,
RA Szpakowski G.V., Ussery D., Barrell B.G., Nurse P.;
RT "The genome sequence of Schizosaccharomyces pombe.";
RL Nature 415:871-880(2002).
RN SEQUENCE OF 136-543 FROM N.A.
RP Kominami K., Toda T.;
RA Submitted (AUG-1998) to the EMBL/GenBank/DBJ databases.
RL -!- SIMILARITY: Belongs to the cullin family.
CC -!- CAUTION: Ref.2 sequence differs from that shown due to a
CC frameshift in position 513.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; Z54142; CRA90847.1; -.
CC EMBL; AB017028; BAA32519.1; ALT_FRAME.
CC PIR; S62405; S62405.
CC PIR; T38359; T38359.
CC GeneDB SPombe; SPAC24H6.03; -.
CC InterPro; IPR001373; Cullin.
CC Pfam; PF00888; Cullin; 1.
CC SMART; SM00182; CULLIN; 1.
CC PROSITE; PS01256; CULLIN_1; 1.
CC PROSITE; PS00069; CULLIN_2; 2.
CC CONFLICT 476 476 R -> RYALIVFTVFNTR (IN REF. 2).
FT CONFLICT 476 476 R -> RYALIVFTVFNTR (IN REF. 2).
SQ SEQUENCE 785 AA; 90511 MW; 0830F7A857DF6CE7 CRC64;

Query Match 2.9%; Score 7; DB 1; Length 785;
Best Local Similarity 100.0%; Pred. No. 61;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 82 ETTSQL 88
Db 353 ETTSQL 359

RESULT 27
ROR1_HUMAN
ID ROR1_HUMAN STANDARD; PRT; 937 AA.
AC Q01973; Q92776;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Tyrosine-protein kinase transmembrane receptor ROR1 precursor
DE (EC 2.7.1.112) (Neurotrophic tyrosine kinase, receptor-related 1).
DE ROR1 OR NTRK1.
GN ROR1 sapiens (Human).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN SEQUENCE FROM N.A. (ISOFORM LONG).
RP MEDLINE=93100347; PubMed=1334494;
```

DR	PROSITE; PS50835; IG LIKE, 1.
DR	PROSITE; PS00021; KRINGLE_1; 1.
DR	PROSITE; PS00070; KRINGLE_2; 1.
DR	PROSITE; PS00107; PROTEIN KINASE ATP; FALSE_NEG.
DR	PROSITE; PS00117; PROTEIN KINASE DOM; 1.
DR	PROSITE; PS00109; PROTEIN KINASE TYR; 1.
KW	Transferase; Tyrosine-protein kinase; ATP-binding; Receptor;
KW	Transmembrane; signal; Glycoprotein; Kringle; Phosphorylation;
KW	Immunoglobulin domain; Alternative splicing.
FT	SIGNAL 1 29
FT	CHAIN 30 937
FT	TROSYNE-PROTEIN KINASE TRANSMEMBRANE
FT	RECEPTOR RORI.
FT	EXTRACELLULAR (POTENTIAL).
FT	POTENTIAL.
FT	CYTOPLASMIC (POTENTIAL).
FT	IG-LIKE C2-TYPE.
FT	FZ.
FT	KRINGLE.
FT	PROTEIN KINASE.
FT	SER/THR-RICH.
FT	PRO-RICH.
FT	SER/THR-RICH.
FT	NP BIND 479 487
FT	ATP (BY SIMILARITY).
FT	BY SIMILARITY.
FT	PHOSPHORYLATION (AUTO-) (BY SIMILARITY)
FT	BY SIMILARITY.
FT	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	Missing (in isoform Short).
FT	/FTid=VSP_005008.
FT	SEQUENCE 937 AA; 104312 MW; OD0694DBF2984773 CRC64;
QY	Query Match 2.9%; Score 7; DB 1; Length 937;
DB	Best Local Similarity 100.0%; Pred. No. 70;
	Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps
QY	46 HSYCRNP 52
DB	359 HSYCRNP 365
RESULT 28	
RORI_MOUSE	
ID AC Q92I39; STANDARD; PRT; 937 AA.	
DT 16-OCT-2001 (Rel. 40, Created)	
DT 16-OCT-2001 (Rel. 40, Last sequence update)	
DT 10-OCT-2003 (Rel. 42, Last annotation update)	
DE Tyrosine-protein kinase transmembrane receptor RORI precursor	
DE (EC 2.7.1.112) (Neurotrophic tyrosine kinase, receptor-related 1)	
DE (mRORI).	
GN RORI OR NTRKRI.	
GC Mus musculus (Mouse).	
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
OX Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.	
OX NCBI_TaxID=10090;	
RN [1]	
RP SEQUENCE FROM N.A.	
RX MEDLINE=99248426; PubMed=10231392;	
RA Oishi I., Takeuchi S., Hashimoto R., Nagabuku A., Ueda T., Liu Z.J.	
RA Hatta T., Akita S., Matsuda Y., Yamana H., Orani H., Minami Y.;	
RT "Spatio-temporally regulated expression of receptor tyrosine kinases	
RT mRori, mRor2, during mouse development: implications in development	
RT and function of the nervous system.";	
RL Genes Cells 4:41-56(1999)	
CC -!- FUNCTION: Tyrosine-protein kinase receptor whose role is not yet	
CC clear.	
CC -!- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein	
CC tyrosine phosphate.	
CC -!- SUBCELLULAR LOCATION: Type I membrane protein.	



```

Query Match          2.9%; Score 7; DB 1; Length 937;
Best Local Similarity 100.0%; Pred. No. 70;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
RA Svirakas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
RA Wang Z.-Y., Wasserman D.A., Weinstock G.M., Weissenbach J.,  
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Smith H.O.,  
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
RT "The genome sequence of *Drosophila melanogaster*";  
RL Science 287:2185-2195(2000).  
RN [5]  
RP REVISIONS.  
RX MEDLINE=22426069; PubMed=12537572;  
RA Misra S., Crosby M.A., Mungall C.J., Matthews B.B., Campbell K.S.,  
RA Hradecky P., Huang Y., Kaminker J.S., Millburn G.H., Prochuk S.E.,  
RA Smith C.D., Tupy J.L., Whitfield E.J., Bayraktaroglu L., Berman B.P.,  
RA Bettencourt B.R., Celisner S.E., de Grey A.D.N.J., Drysdale R.A.,  
RA Harris N.L., Richter J., Russo S., Schroeder A.J., Shu S.Q.,  
RA Stapleton M., Yamada C., Ashburner M., Gelbart W.M., Rubin G.M.,  
RA Lewis S.E.;  
RT "Annotation of the *Drosophila melanogaster* euchromatic genome: a  
RT systematic review";  
RL Genome Biol. 3:RESEARCH0083.1-RESEARCH0083.22(2002).  
CC -!- FUNCTION: Part of the AP-3 complex, an adaptor-related complex  
CC which is not clathrin-associated. The complex is associated with  
CC the Golgi region as well as more peripheral structures. It  
CC facilitates the budding of vesicles from the Golgi membrane and  
CC may be directly involved in trafficking to lysosomes (By  
CC similarity).  
CC -!- FUNCTION: May be a coat protein involved in the formation of  
CC specialized structures like pigment granules.  
CC -!- SUBUNIT: Assembly protein complex 3 (AP-3) is a heterotrimer  
CC composed of two large chains (delta and beta3), a medium chain  
CC (mu3) and a small chain (sigma3).  
CC -!- SUBCELLULAR LOCATION: Component of the coat surrounding the  
CC cytoplasmic face of coated vesicles located at the Golgi complex  
CC (By similarity).  
CC -!- SIMILARITY: Belongs to the adaptor complexes large subunit  
CC family.  
CC -!- CAUTION: Sequences in Ref.2 and Ref.3 contain intronic sequences  
CC and are incomplete at 5' and 3' ends. Sequences extensively differ  
CC from that shown at positions 1-269, 546, and 840-1034.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; AF002164; AAC14585.1; -  
CC EMBL; AF044287; AAC01743.1; ALT\_SEQ.  
CC EMBL; U33351; AAB97618.1; ALT\_SEQ.  
CC EMBL; AE003493; AAF48307.2; -  
CC FlyBase; FBgn001087; g.  
CC GO; GO:0030123; C:AP-3 adaptor complex; NAS.  
CC GO; GO:0005783; C:endoplasmic reticulum; IDA.  
CC GO; GO:0005795; C:Golgi stack; IDA.  
CC GO; GO:0005798; C:Golgi vesicle; NAS.  
CC GO; GO:0008057; P:eye pigment granule morphogenesis (sensu Dr. .; IMP.  
CC GO; GO:0006895; P:Golgi to endosome transport; NAS.  
CC GO; GO:0006727; P:omochromosome biosynthesis; IMP.  
CC InterPro; IPR002553; Adaptin\_N.  
CC InterPro; IPR008938; ARM.  
CC Pfam; PF01602; Adaptin\_N; 1.  
CC Golgi stack; Protein transport; Transport.  
KW DOMAIN 767 785 LYS-RICH  
FT CONFLICT 395 395 D -> T (IN REF. 2).  
FT CONFLICT 423 423 L -> V (IN REF. 2).  
FT CONFLICT 442 445 QLLD -> RTTY (IN REF. 2).  
FT

FT CONFLICT 461 465 MTNLL -> IDQSA (IN REF. 2).  
FT CONFLICT 694 695 QR -> GO (IN REF. 2).  
FT CONFLICT 701 701 E -> D (IN REF. 1).  
FT CONFLICT 869 869 L -> S (IN REF. 4).  
FT CONFLICT 910 910 P -> S (IN REF. 4).  
SQ SEQUENCE 1034 AA; 114845 MW; ECE1B7A34DCSF8F1 CRC64;  
Query Match 2.9%; Score 7; DB 1; Length 1034;  
Best Local Similarity 100.0%; Pred. No. 77;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 123 AAQVPVI 129  
Db 906 AAQVPVI 912  
RESULT 30  
C9CA BACTO  
ID C9CA BACTO STANDARD; PRT; 1157 AA.  
AC Q45733;  
DT 30-MAY-2000 (Rel. 39, Created)  
DT 30-MAY-2000 (Rel. 39, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Pesticidal crystal protein cry3ca (Insecticidal delta-endotoxin  
DE CryXC(a)) (Crystalline entomocidal protoxin) (130 kDa crystal  
DE protein).  
GN CRY9CA OR CRYXC(A).  
OS *Bacillus thuringiensis* (subsp. *tolworthi*).  
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.  
OX NCBI\_TaxID=1442;  
RN [1]  
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.  
RC STRAIN=BTS02618A;  
RX MEDLINE=96141404; PubMed=8572715;  
RA Lambert B., Buyse L., Decock C., Jansens S., Piens C., Saey B.,  
RA Seurinck J., Van Audenhove K., Van Rie J., Van Vliet A., Peferoen M.;  
RT "A *Bacillus thuringiensis* insecticidal crystal protein with a high  
RT activity against members of the family Noctuidae".  
RL Appl. Environ. Microbiol. 62:80-86(1996).  
CC -!- FUNCTION: PROMOTES COLLOIDSMOTIC LYSIS BY BINDING TO THE MIDGUT  
CC EPITHELIAL CELLS OF LEPIDOPTERA LARVAE. HAS A FAIRLY BROAD  
CC SPECTRUM OF ACTIVITY AGAINST MEMBERS OF THE FAMILIES PYRALIDAE,  
CC PLUTELLIDAE, SPHINGIDAE, AND NOCTUIDAE. IT IS THE FIRST  
CC INSECTICIDAL CRYSTAL PROTEIN WITH ACTIVITY AGAINST CUTWORMS. NO  
CC ACTIVITY IS OBSERVED AGAINST SOME BEETLES, SUCH AS COLORADO POTATO  
CC BEETLE.  
CC -!- DEVELOPMENTAL STAGE: The crystal protein is produced during  
CC sporulation and is accumulated both as an inclusion and as part of  
CC the spore coat.  
CC -!- MISCELLANEOUS: Toxic segment of the protein is located in the N-  
CC terminus.  
CC -!- SIMILARITY: Belongs to the delta endotoxin family.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; Z37527; CAA85764.1; -  
CC PIR; A59350; S49247.  
CC HSP; P01130; 1DLC.  
CC InterPro; IPR001178; Endotoxin.  
CC InterPro; IPR005638; endotoxin\_C.  
CC InterPro; IPR005639; endotoxin\_N.  
CC InterPro; IPR008979; Gal\_bind\_like.  
CC Pfam; PF00555; endotoxin\_1.  
CC Pfam; PF03944; endotoxin\_C; 1.  
CC Pfam; PF03945; endotoxin\_N; 1.  
CC Toxin; Sporulation.  
KW SEQUENCE 1157 AA; 129775 MW; C364391EF7FDFB8A CRC64;  
SQ

```

Query Match          2.9%; Score 7; DB 1; Length 1157;
Best Local Similarity 100.0%; Pred. No. 84;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 36 ASAPVSG 42
Db 523 ASAPVSG 529

RESULT 31
POLG_KUNJM
ID POLG_KUNJM STANDARD; PRT; 3433 AA.
AC P14335; Q82983;
DT 01-JAN-1990 (Rel. 13, Last sequence update)
DT 01-JAN-1990 (Rel. 13, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Genome polyprotein [Contains: Capsid protein C (Core protein); Matrix
DE protein (Envelope protein M); Major envelope protein E; Nonstructural
DE proteins NS1, NS2, NS4A and NS4B; Protease/helicase (EC 3.4.21.98)
DE (NS3); RNA-directed RNA polymerase (EC 2.7.7.48) (NS5)].
OS Kunjin virus (strain MRM61C).
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Flavivirus.
OX NCBI_TaxID=11078;
RN [1]_TaxID=11078;
SEQUENCE FROM N.A.
RP MEDLINE=88089524; PubMed=2826659;
RA Coia G., Parker M.D., Speight G., Byrne M.E., Westaway E.G.;
RT "Nucleotide and complete amino acid sequences of Kunjin virus:
RT definitive gene order and characteristics of the virus-specified
RT proteins.";
RL J. Gen. Virol. 69:1-21(1988).
CC -!- FUNCTION: The small proteins NS2A, NS2B, NS4A and NS4B are
CC hydrophobic, suggesting a possible membrane-related function.
CC NS3 and NS5 may play a role in the viral RNA replication.
CC -!- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral
CC precursor polyprotein, commonly with Asp or Glu in the P6
CC position, Cys or Thr in P1 and Ser or Ala in P1'.
CC -!- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate +
CC {RNA} (N).
CC -!- SUBUNIT: The virion of this virus is a nucleocapsid covered by a
CC lipoprotein envelope. The envelope consists of two proteins:
CC protein M and glycoprotein E. The nucleocapsid is a complex of
CC protein C and mRNA.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC -----
CC EMBL; D00246; BAA00176.1; -
CC PIR; A28697; GNWVKV.
CC KSP; P14336; 1SVB.
CC
CC MEROPS; S07.001; -
CC
CC InterPro; IPR009003; Cys_Ser_trypsin.
CC InterPro; IPR001410; DEAD.
CC InterPro; IPR001122; Flavi_capsidC.
CC InterPro; IPR000336; Flavi_glycoprote.
CC InterPro; IPR000699; Flavi M.
CC InterPro; IPR001157; Flavi NS1.
CC InterPro; IPR000752; Flavi NS2A.
CC InterPro; IPR000487; Flavi NS2B.
CC InterPro; IPR000404; Flavi NS4A.
CC InterPro; IPR001528; Flavi NS4B.
CC InterPro; IPR000208; Flavi NS5.
CC InterPro; IPR002535; Flavi NS5.
CC InterPro; IPR001650; Helicase_C.
CC InterPro; IPR007110; IG-like.
CC InterPro; IPR001850; Peptidase_S7.

```

```

DR InterPro; IPR007095; RNA_pol_DS_PS.
DR InterPro; IPR007094; RNA_pol_PSVir.
DR InterPro; IPR002877; RmJ_PtsJ.
DR Pfam; PF01003; Flavi_capsid; 1.
DR Pfam; PF02832; Flavi_glycop_C; 1.
DR Pfam; PF00869; Flavi_glycop_C; 1.
DR Pfam; PF00949; Flavi_helicase; 1.
DR Pfam; PF01004; Flavi_M; 1.
DR Pfam; PF00948; Flavi_NS1; 1.
DR Pfam; PF01005; Flavi_NS2A; 1.
DR Pfam; PF01002; Flavi_NS2B; 1.
DR Pfam; PF01350; Flavi_NS4A; 1.
DR Pfam; PF01349; Flavi_NS4B; 1.
DR Pfam; PF00972; Flavi_NS5; 1.
DR Pfam; PF01570; Flavi_propep; 1.
DR Pfam; PF01728; PtsJ; 1.
DR Pfam; PF00211; helicase_C; 1.
DR ProDom; PF001556; Flavi_glycoprote; 1.
DR ProDom; PF001496; Flavi_NS1; 1.
DR SMART; SMO0487; DEXDC; 1.
DR SMART; SMO0490; HELIC; 1.
DR PROSITE; PS00690; DEAH ATP HELICASE; FALSE NEG.
KW Polyprotein; Glycoprotein; Transferase; RNA-directed RNA polymerase;
KW Core protein; Coat protein; Envelope protein; Hydrolase; Helicase;
KW ATP-binding; Transmembrane; Nonstructural protein
FW INIT_MET 1 1
FW INIT_MET 1 1
FW CHAIN 1 123
FW CHAIN 124 215
FW CHAIN 216 290
FW CHAIN 291 791
FW CHAIN 792 1143
FW CHAIN 1144 1374
FW CHAIN 1375 1505
FW CHAIN 1506 2124
FW CHAIN 2125 2273
FW CHAIN 2274 2528
FW CHAIN 2529 3433
FW DOMAIN 388 401
FW NP_BIND 1699 1706
FW SITE 1790 1793
FW DISULFID 293 320
FW DISULFID 350 406
FW DISULFID 364 395
FW DISULFID 382 411
FW DISULFID 480 578
FW DISULFID 595 626
FW CARBOHYD 138 138
FW CARBOHYD 921 921
FW CARBOHYD 966 966
FW CARBOHYD 998 998
FW SEQUENCE 3433 AA; 381363 MW; E54B889A7D040B99 CRC64;

Query Match          2.9%; Score 7; DB 1; Length 3433;
Best Local Similarity 100.0%; Pred. No. 2.1e-02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 21 PGLRCLN 27
Db 1307 PGLRCLN 1313

RESULT 32
LAVA_EMENI
ID LAVA_EMENI STANDARD; PRT; 41 AA.
AC P38095;
DT 01-OCT-1994 (Rel. 30, Created)
DT 01-OCT-1994 (Rel. 30, Last sequence update)
DT 01-OCT-1994 (Rel. 30, Last annotation update)
DE Lactam utilization protein LAMA (Fragment).
GN LAMA.
OS Emericella nidulans (Aspergillus nidulans).
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;

```

```

OC Eurotiales; Trichocomaceae; Emericella.
OX NCBI_TaxID=162425;
RN [1]_SEQUENCE FROM N.A.
RP MEDLINE=92107186; PubMed=1729609;
RA Richardson I.B., Katz M.E., Hynes M.J.;
RX "Molecular characterization of the lam locus and sequences involved
RT in regulation by the AmrR protein of Aspergillus nidulans.";
RL Mol. Cell. Biol. 12:337-346(1992).
CC -!- FUNCTION: Involved in the utilization of lactams such as 2-
CC pyroglutamine.
CC -!- INDUCTION: By beta-alanine.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; M77283; AAA33312.1; ALT_SEQ.
DR PIR; A42064; A42064.
FT NON_TER 41
SQ SEQUENCE 41 AA; 4458 MW; D9BE04A383A78FF2 CRC64;

Query Match 2.5%; Score 6; DB 1; Length 41;
Best Local Similarity 100.0%; Pred. No. 53;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 17 TSPAPG 22
DB 33 TSPAPG 38

RESULT 33
Y185_METJA
ID Y185_METJA STANDARD; PRT; 49 AA.
AC Q5764;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical protein M70185.
GN M70185.
OS Methanococcus jannaschii.
OC Archaea; Euryarchaeota; Methanococci; Methanococcales;
OC Methanocaldococcaceae; Methanocaldococcus.
OX NCBI_TaxID=2190;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=JAL-1 / DSM 2661 / ATCC 43067;
MEDLINE=9633799; PubMed=868087;
RA Bult C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,
RA Sutton G.G., Blake J.A., FitzGerald L.M., Clayton R.A., Gocayne J.D.,
RA Kerlavage A.R., Dougherty B.A., Tomb J.-P., Adams M.D., Reich C.I.,
RA Overbeek R., Kirkness E.F., Weinstock K.G., Merrick J.M., Glodek A.,
RA Scott J.L., Geoghegan N.S.M., Weidman J.F., Fuhmann J.L., Nguyen D.,
RA Utterback T.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,
RA Cotton M.D., Roberts K.M., Hurst M.A., Kaine B.P., Borodovsky M.,
RA Klenk H.-P., Fraser C.M., Smith H.O., Woese C.R., Venter J.C.;
RT "Complete genome sequence of the methanogenic archaeon, Methanococcus
RT jannaschii";
RL Science 273:1058-1073 (1996).
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; U67475; AAB98182.1;
DR

```

```

DR PIR; B64323; B64323.
DR TIGR; M00185; -.
KW Hypothetical protein; Transmembrane; Complete proteome.
FT TRANSMEM 22
SQ SEQUENCE 49 AA; 5332 MW; D1594F2AAD47CEF4 CRC64;

Query Match 2.5%; Score 6; DB 1; Length 49;
Best Local Similarity 100.0%; Pred. No. 61;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 159 IIIAIG 164
DB 30 IIIAIG 35

RESULT 34.
BD02_RAT
ID BD02_RAT STANDARD; PRT; 63 AA.
AC O88514;
DT 15-JUL-1999 (Rel. 38, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Beta-defensin 2 precursor (BD-2) (RBD-2).
GN DEFB2.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=Wistar;
MEDLINE=9938683; PubMed=10456937;
RA Jia H.P., Mills J.N., Barahmand-Pour F., Nishimura D.,
RA Mallampalli R.K., Wang G., Wiles K., Tack B.F., Bevins C.L.,
RA McCray P.B. Jr.;
RT "Molecular cloning and characterization of rat genes encoding
RT homologues of human beta-defensins.";
RL Infect. Immun. 67:4827-4833(1999).
CC -!- FUNCTION: Has bactericidal activity (By similarity).
CC -!- SUBCELLULAR LOCATION: Secreted (Potential).
CC -!- TISSUE SPECIFICITY: Highly expressed in lung.
CC -!- SIMILARITY: Belongs to the beta-defensin family. LAP/TAP
CC subfamily.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AF068861; AAC28072.1; -.
DR HSSP; O15263; IFD3.
DR InterPro; IPR001855; Defensin_beta.
DR InterPro; IPR006080; Defensin_mammal.
DR Pfam; PF00711; Defensin_beta; 1.
DR SMART; SM00048; DEFSN; 1.
KW Antibiotic; Signal.
FT SIGNAL 1
FT PROPEP 21 ?
FT CHAIN 22 63
FT DISULFID 31 59
FT DISULFID 38 52
FT DISULFID 42 60
FT SEQUENCE 63 AA; 6946 MW; 826099DE2144ACF4 CRC64;

Query Match 2.5%; Score 6; DB 1; Length 63;
Best Local Similarity 100.0%; Pred. No. 76;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 198 PLSAFT 203
DB 198 PLSAFT 203

```

```

Db      17 PLSAFT 22

RESULT 35
CSPA_PSEAE STANDARD; PRT; 69 AA.
AC P95459;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Major cold shock protein cspa.
GN CSPA OR PA3265.
OS Pseudomonas aeruginosa.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
OC Pseudomonadaceae; Pseudomonas.
OX NCBI_TaxID=287;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 15692 / PA01;
RA Francis K.P., Stewart G.S.A.B.;
RT "Gene duplication: a mechanism for the evolution of bacterial major
RL cold shock protein families.";
RL Submitted (JAN-1997) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 15692 / PA01;
RX MEDLINE=20437337; PubMed=10984043;
RA Hickey M.J., Brinkman F.S.L., Hufnagle W.O., Kowalik D.J., Lagrou M.,
RA Garber R.L., Goltzy L., Tolentino E., Westbrook-Wadman S., Yuan Y.,
RA Brody L.L., Coulter S.N., Folger K.R., Kas A., Larbig K., Lim R.M.,
RA Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,
RA Reizer J., Saier M.H., Hancock R.E.W., Lory S., Olson M.V.;
RA "Complete genome sequence of Pseudomonas aeruginosa PA01, an
RT opportunistic pathogen.";
RL Nature 406:959-964 (2000).
CC -1- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -1- INDUCTION: In response to low temperature.
CC -1- SIMILARITY: Belongs to the cold-shock domain (CSD) family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; U82822; AAB40922.1; --
CC EMBL; AB004749; AAG06654.1; --
CC F01; H83236; H83236.
CC HSP; P32081; 1CSP.
CC InterPro; IPR002059; Cold shock.
CC InterPro; IPR008994; Nucleic_acid_OR.
CC Pfam; PF00313; CSD; 1.
CC PRINTS; PR00050; COLDSHOCK.
CC ProDom; PD000621; Cold_shock; 1.
CC SMART; SM00357; CSP; 1.
CC PROSITE; PS00352; COLD_SHOCK; 1.
CC Transcription regulation; DNA-binding; Activator; Multigene family;
CC Complete proteome.
FT DOMAIN 7 66
SQ SEQUENCE 69 AA; 7606 MW; 993225127C41AE43 CRC64;

Query Match 2.5%; Score 6; DB 1; Length 69;
Best Local Similarity 100.0%; Pred.No. 82;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 104 ADEVQV 109
Db 63 ADEVQV 68
|||||
OC Bacteria; Actinobacteria; Actinomycetales;
OC Corynebacteriaceae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1769;
RN [1]

RESULT 36
PYSI_SYNEL STANDARD; PRT; 78 AA.
AC P50035;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Phycobilisome 8.9 kDa linker polypeptide, phycocyanin-associated, rod
DE (rod capping linker protein).
DE CPD OR ISR1960.
GN Synechococcus elongatus (Thermosynechococcus elongatus).
OC Bacteria; Cyanobacteria; Chroococcales; Synechococcus.
OX NCBI_TaxID=32046;
RN [1]
RP SEQUENCE FROM N.A.
RA Shimazu T., Soga M., Hirano M., Kato S.;
RT "Cloning and sequencing of the phycocyanin operon from the
RT thermophilic cyanobacterium Synechococcus elongatus.";
RL Submitted (SEP-1992) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=BP-1;
RX MEDLINE=2225144; PubMed=12240834;
RA Nakamura Y., Kaneko T., Sato S., Ikeuchi M., Kato H., Sasamoto S.,
RA Watanabe A., Iriguchi M., Kawashima K., Kimura T., Kishida Y.,
RA Kiyokawa C., Kohara M., Matsumoto M., Matsumoto A., Nakazaki N.,
RA Shimo S., Sugimoto M., Takeuchi C., Yamada M., Tabata S.;
RA "Complete genome structure of the thermophilic cyanobacterium
RT Thermosynechococcus elongatus BP-1.";
RL DNA Res. 9:123-130 (2002).
CC -1- SIMILARITY: Belongs to the phycobilisome linker protein family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; D13173; BAA02458.1; --
CC EMBL; AP005375; BAC09512.1; --
CC InterPro; IPR008213; CpcD-like.
CC InterPro; IPR008214; CpcD-like_C.
CC Pfam; PF01383; CpcD; 1.
CC ProDom; PD002828; CpcD-like_C; 1.
CC Phycobilisome; Photosynthesis; Complete proteome.
KW Phycobilisome; 78 AA; 8672 MW; E529978375f30247 CRC64;
SQ SEQUENCE 78 AA; 8672 MW; E529978375f30247 CRC64;

Query Match 2.5%; Score 6; DB 1; Length 78;
Best Local Similarity 100.0%; Pred.No. 91;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 191 EMQKIT 196
Db 58 EMQKIT 63
|||||
ATPPL MYCLE STANDARD; PRT; 81 AA.
ID ATPPL_MYCLE
AC P45828;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE ATP synthase C chain (EC 3.6.3.14) (lipid-binding protein).
DE (Dicyclohexylcarbodiimide-binding protein).
DE ATP OR ML1140.
GN Mycobacterium leprae.
OS Mycobacterium leprae.
OC Bacteria; Actinobacteria; Actinomycetales;
OC Corynebacteriaceae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1769;
RN [1]

```

RP SEQUENCE FROM N.A.  
RA Smith D.R., Robison K.;  
RL Submitted (SEP-1994) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RC STRAIN=TN;  
RX MEDLINE=21128732; PubMed=11234002;  
RA Cole S.T., Bigmeier K., Parkhill J., James K.D., Thomson N.R.,  
RA Wheeler P.R., Honore N., Garnier T., Churcher C., Harris D.,  
RA Mungall K., Basham D., Brown D., Chillingworth T., Connor R.,  
RA Davies R.M., Devlin K., Duthoy S., Feltwell T., Fraser A., Hamlin N.,  
RA Holroyd S., Hornsby T., Jagels K., Lacroix C., Maclean J., Moule S.,  
RA Murphy L., Oliver K., Quail M.A., Rajandream M.A., Rutherford K.M.,  
RA Rutter S., Seeger K., Simon S., Simmonds M., Skelton J., Squares R.,  
RA Squares S., Stevens K., Taylor K., Whitehead S., Woodward J.R.,  
RA Barrett B.G.;  
RL "Massive gene decay in the leprosy bacillus";  
RT Nature 409:1007-1011(2001).  
CC -!- FUNCTION: This is one of the three chains of the nonenzymatic  
CC component (CF(0) subunit) of the ATPase complex.  
CC -!- CATALYTIC ACTIVITY: ATP + H(2)O + H(+) (in) = ADP + phosphate +  
CC H(+) (out).  
CC -!- SUBUNIT: F-type ATPases have 2 components, CF(1) - the catalytic  
CC core - and CF(0) - the membrane proton channel. CF(1) has five  
CC subunits: alpha(3), beta(3), gamma(1), delta(1), epsilon(1). CF(0)  
CC has three main subunits: a, b and c.  
CC -!- MISCELLANEOUS: Dicyclohexylcarbodiimide (DCCD) inhibits ATPase.  
CC -!- SIMILARITY: Belongs to the ATPase C chain family.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC -----  
CC EMBL; U15186; AAA63107.1; -;  
CC EMBL; AL583920; CAC31521.1; -;  
CC PIR; T09979; T09979;  
CC Leproma; M11140; -;  
CC InterPro; IPR005953; ATP synth C;  
CC InterPro; IPR002379; ATPase\_Csub.  
CC Pfam; PF00137; ATP-synt C; 1.  
CC PRINTS; PR00124; ATPASEC.  
CC TIGRfams; TIGR01260; ATP synt c; 1.  
CC PROSITE; PS00605; ATPase\_C\_1;  
CC Hydrogen ion transport; CF(0); Transmembrane; Lipid-binding;  
CC Complete proteome.  
CC TRANSMEM 5 25  
CC TRANSMEM 57 77  
CC BINDING 61 61  
CC DICYCLOHEXYLCARBODIIMIDE (BY SIMILARITY).  
CC SEQUENCE 81 AA; 8205 MW; E5456207EE9EC13 CRC64;  
Query Match 2.5%; Score 6; DB 1; Length 81;  
Best Local Similarity 100.0%; Pred. No. 94;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 162 AIGAGI 167  
Db 21 AIGAGI 26  
RESULT 38  
ATPL MYCTU STANDARD; PRT; 81 AA.  
AC Q10598;  
DT 01-OCT-1996 (Rel. 34, Created)  
DT 01-OCT-1996 (Rel. 34, Last sequence update)  
DT 01-OCT-2003 (Rel. 42, Last annotation update)  
DE ATP synthase C chain (EC 3.6.3.14) (lipid-binding protein)  
DE (Dicyclohexylcarbodiimide-binding protein).

GN ATPB OR RV1305 OR MT1345 OR MTCY373.25 OR MB1337.  
OS Mycobacterium tuberculosis, and  
OS Bacteria; Actinobacteridae; Actinobacteriales;  
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.  
OX NCBI\_TaxID=1773, 1765;  
RN [1]  
RC SEQUENCE FROM N.A.  
RX MEDLINE=21128732; PubMed=11234002;  
RA Cole S.T., Bigmeier K., Parkhill J., James K.D., Thomson N.R.,  
RA Wheeler P.R., Honore N., Garnier T., Churcher C., Harris D.,  
RA Mungall K., Basham D., Brown D., Chillingworth T., Connor R.,  
RA Davies R.M., Devlin K., Duthoy S., Feltwell T., Fraser A., Hamlin N.,  
RA Holroyd S., Hornsby T., Jagels K., Lacroix C., Maclean J., Moule S.,  
RA Murphy L., Oliver K., Quail M.A., Rajandream M.A., Rutherford K.M.,  
RA Rutter S., Seeger K., Simon S., Simmonds M., Skelton J., Squares R.,  
RA Squares S., Stevens K., Taylor K., Whitehead S., Woodward J.R.,  
RA Barrett B.G.;  
RL "Massive gene decay in the leprosy bacillus";  
RT Nature 409:1007-1011(2001).  
CC -!- FUNCTION: This is one of the three chains of the nonenzymatic  
CC component (CF(0) subunit) of the ATPase complex.  
CC -!- CATALYTIC ACTIVITY: ATP + H(2)O + H(+) (in) = ADP + phosphate +  
CC H(+) (out).  
CC -!- SUBUNIT: F-type ATPases have 2 components, CF(1) - the catalytic  
CC core - and CF(0) - the membrane proton channel. CF(1) has five  
CC subunits: alpha(3), beta(3), gamma(1), delta(1), epsilon(1). CF(0)  
CC has three main subunits: a, b and c (By similarity).  
CC -!- MISCELLANEOUS: Dicyclohexylcarbodiimide (DCCD) inhibits ATPase.  
CC -!- SIMILARITY: Belongs to the ATPase C chain family.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC -----  
CC EMBL; 273419; CAA97738.1; -;  
CC EMBL; AE007008; AAK45607.1; -;  
CC EMBL; EX248335; CND94198.1; -;  
CC FIR; E70774; E70774;  
CC TIGR; MT1345; -;  
CC Tuberculin; RV1305; -;  
CC InterPro; IPR005953; ATP synth C;  
CC InterPro; IPR002379; ATPase\_Csub.  
CC Pfam; PF00137; ATP-synt C; 1.  
CC PRINTS; PR00124; ATPASEC.  
CC TIGRfams; TIGR01260; ATP synt c; 1.  
CC PROSITE; PS00605; ATPase\_C\_1;  
CC Hydrogen ion transport; CF(0); Transmembrane; Lipid-binding;  
CC Complete proteome.  
CC TRANSMEM 5 25  
CC TRANSMEM 57 77  
CC BINDING 61 61  
CC DICYCLOHEXYLCARBODIIMIDE (BY SIMILARITY).  
CC SEQUENCE 81 AA; 8205 MW; E5456207EE9EC13 CRC64;  
Query Match 2.5%; Score 6; DB 1; Length 81;  
Best Local Similarity 100.0%; Pred. No. 94;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 162 AIGAGI 167  
Db 21 AIGAGI 26  
RESULT 38  
ATPL MYCTU STANDARD; PRT; 81 AA.  
AC Q10598;  
DT 01-OCT-1996 (Rel. 34, Created)  
DT 01-OCT-1996 (Rel. 34, Last sequence update)  
DT 01-OCT-2003 (Rel. 42, Last annotation update)  
DE ATP synthase C chain (EC 3.6.3.14) (lipid-binding protein)  
DE (Dicyclohexylcarbodiimide-binding protein).

DR TIGR01260; ATP\_synt\_c; 1.  
DR PROSITE; PS00605; ATPASE\_C; 1.  
KW Hydrogen ion transport; CF(0); Transmembrane; Lipid-binding;  
Complete proteome.  
FT TRANSMEM 5 25 POTENTIAL.  
FT TRANSMEM 57 77 POTENTIAL.  
FT BINDING 61 61 DICYCLOHEXYLCARBODIIMIDE (BY SIMILARITY).  
FT SEQUENCE 81 AA; 8055 MW; 36303496CDDF52CC CRC64;  
SQ  
Query Match 2.5%; Score 6; DB 1; Length 81;  
Best Local Similarity 100.0%; Pred.No. 94;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 162 AIGAGI 167  
|||||  
DB 21 AIGAGI 26  
RESULT 39  
RS17\_CHLNU  
ID RS17\_CHLNU STANDARD; PRT; 83 AA.  
AC Q9JUM3;  
DT 16-OCT-2001 (Rel. 40, Created)  
DT 16-OCT-2001 (Rel. 40, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE 30S ribosomal protein S17.  
GN RPSQ OR TC0806.  
OS Chlamydia muridarum.  
OC Bacteria; Chlamydiae; Chlamydiales; Chlamydiaceae; Chlamydia.  
OX NCBI\_TaxID=83560;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=MoPn / Ni99;  
RX MEDLINE=20150255; PubMed=10684935;  
RA Read T.D., Brunham R.C., Shen C., Gill S.R., Heidelberg J.F.,  
White O., Hickey E.K., Peterson J., Utterback T., Barry K., Bass S.,  
Linher K., Weidman J., Khouri H., Craven B., Bowman C., Dodson R.,  
Rauw G., Nelson W., DeBoy R., Kolonay J., McClarty G., Salzberg S.L.,  
Eisen J., Fraser C.W.;  
RA "Genome sequences of Chlamydia trachomatis MoPn and Chlamydia  
pneumoniae AR39.";  
RT Nucleic Acids Res. 28:1397-1406(2000).  
RL Nucleic Acids Res. 28:1397-1406(2000).  
CC -!- FUNCTION: Protein S17 binds specifically to the 5' end of 16S  
ribosomal RNA (By similarity).  
CC -!- SIMILARITY: Belongs to the S17P family of ribosomal proteins.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC  
CC EMBL; AF002347; AAF39609.1; -.  
DR EMBL; AF002347; AAF39609.1; -.  
DR EMBL; F81664; F81664.  
DR HSSP; P23828; LRIP.  
DR TIGR; TC0806;  
DR InterPro; IPR008994; Nucleic acid OB.  
DR InterPro; IPR00266; Ribosomal S17.  
DR Pfam; PF00366; Ribosomal\_S17; 1.  
DR PRINTS; PR00973; RIBOSOMALS17.  
DR ProDom; PD001295; Ribosomal\_S17; 1.  
DR PROSITE; PS00056; RIBOSOMAL\_S17; 1.  
KW Ribosomal protein; rRNA-binding; Complete proteome.  
SQ SEQUENCE 83 AA; 9703 MW; 446CB5FF64418FD2 CRC64;  
Query Match 2.5%; Score 6; DB 1; Length 83;  
Best Local Similarity 100.0%; Pred.No. 96;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 212 EKTVVV 217  
|||||

DR TIGR01260; ATP\_synt\_c; 1.  
DR PROSITE; PS00605; ATPASE\_C; 1.  
KW Hydrogen ion transport; CF(0); Transmembrane; Lipid-binding;  
Complete proteome.  
FT TRANSMEM 5 25 POTENTIAL.  
FT TRANSMEM 57 77 POTENTIAL.  
FT BINDING 61 61 DICYCLOHEXYLCARBODIIMIDE (BY SIMILARITY).  
FT SEQUENCE 81 AA; 8055 MW; 36303496CDDF52CC CRC64;  
SQ  
Query Match 2.5%; Score 6; DB 1; Length 81;  
Best Local Similarity 100.0%; Pred.No. 94;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 162 AIGAGI 167  
|||||  
DB 21 AIGAGI 26  
RESULT 39  
RS17\_CHLNU  
ID RS17\_CHLNU STANDARD; PRT; 83 AA.  
AC Q9JUM3;  
DT 16-OCT-2001 (Rel. 40, Created)  
DT 16-OCT-2001 (Rel. 40, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE 30S ribosomal protein S17.  
GN RPSQ OR TC0806.  
OS Chlamydia muridarum.  
OC Bacteria; Chlamydiae; Chlamydiales; Chlamydiaceae; Chlamydia.  
OX NCBI\_TaxID=83560;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=MoPn / Ni99;  
RX MEDLINE=20150255; PubMed=10684935;  
RA Read T.D., Brunham R.C., Shen C., Gill S.R., Heidelberg J.F.,  
White O., Hickey E.K., Peterson J., Utterback T., Barry K., Bass S.,  
Linher K., Weidman J., Khouri H., Craven B., Bowman C., Dodson R.,  
Rauw G., Nelson W., DeBoy R., Kolonay J., McClarty G., Salzberg S.L.,  
Eisen J., Fraser C.W.;  
RA "Genome sequences of Chlamydia trachomatis MoPn and Chlamydia  
pneumoniae AR39.";  
RT Nucleic Acids Res. 28:1397-1406(2000).  
RL Nucleic Acids Res. 28:1397-1406(2000).  
CC -!- FUNCTION: Protein S17 binds specifically to the 5' end of 16S  
ribosomal RNA (By similarity).  
CC -!- SIMILARITY: Belongs to the S17P family of ribosomal proteins.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC  
CC EMBL; AF002347; AAF39609.1; -.  
DR EMBL; AF002347; AAF39609.1; -.  
DR EMBL; F81664; F81664.  
DR HSSP; P23828; LRIP.  
DR TIGR; TC0806;  
DR InterPro; IPR008994; Nucleic acid OB.  
DR InterPro; IPR00266; Ribosomal S17.  
DR Pfam; PF00366; Ribosomal\_S17; 1.  
DR PRINTS; PR00973; RIBOSOMALS17.  
DR ProDom; PD001295; Ribosomal\_S17; 1.  
DR PROSITE; PS00056; RIBOSOMAL\_S17; 1.  
KW Ribosomal protein; rRNA-binding; Complete proteome.  
SQ SEQUENCE 83 AA; 9703 MW; 446CB5FF64418FD2 CRC64;  
Query Match 2.5%; Score 6; DB 1; Length 83;  
Best Local Similarity 100.0%; Pred.No. 96;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 212 EKTVVV 217  
|||||

Db 22 EKTVVV 27  
RESULT 40  
RS17\_CHLTR  
ID RS17\_CHLTR STANDARD; PRT; 83 AA.  
AC P28545;  
DT 01-DEC-1992 (Rel. 24, Created)  
DT 01-DEC-1992 (Rel. 24, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE 30S ribosomal protein S17.  
GN RPSQ OR RS17 OR CT519.  
OS Chlamydia trachomatis.  
OC Bacteria; Chlamydiae; Chlamydiales; Chlamydiaceae; Chlamydia.  
OX NCBI\_TaxID=813;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=L2/434/Bu;  
RX MEDLINE=92138612; PubMed=1735714;  
RA Kaul R., Gray G.J., Koehnke N.R., Gu L.J.;  
RT "Cloning and sequence analysis of the Chlamydia trachomatis spc  
ribosomal protein gene cluster.";  
RL J. Bacteriol. 174:1205-1212(1992).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=D/UW-3/Cx;  
RX MEDLINE=9900809; PubMed=9784136;  
RA Stephens R.S., Kaiman S., Lammel C.J., Fan J., Marathe R., Aravind L.,  
Mitchell W.P., Olinger L., Tatusov R.L., Zhao Q., Koonin E.V.,  
Davis R.W.;  
RA "Genome sequence of an obligate intracellular pathogen of humans:  
Chlamydia trachomatis.";  
RT Science 282:754-759(1998).  
RL Science 282:754-759(1998).  
CC -!- FUNCTION: Protein S17 binds specifically to the 5' end of 16S  
ribosomal RNA (By similarity).  
CC -!- SIMILARITY: Belongs to the S17P family of ribosomal proteins.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC  
CC EMBL; M80325; AAA23171.1; -.  
DR EMBL; AE001323; AAC68120.1; -.  
DR EMBL; C42645; C42645.  
DR HSSP; P23828; LRIP.  
DR InterPro; IPR008994; Nucleic acid OB.  
DR InterPro; IPR00266; Ribosomal S17.  
DR Pfam; PF00366; Ribosomal\_S17; 1.  
DR PRINTS; PR00973; RIBOSOMALS17.  
DR ProDom; PD001295; Ribosomal\_S17; 1.  
DR PROSITE; PS00056; RIBOSOMAL\_S17; 1.  
KW Ribosomal protein; rRNA-binding; Complete proteome.  
SQ SEQUENCE 83 AA; 9645 MW; 446CB5FDB4418FD2 CRC64;  
Query Match 2.5%; Score 6; DB 1; Length 83;  
Best Local Similarity 100.0%; Pred.No. 96;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 212 EKTVVV 217  
|||||  
DB 22 EKTVVV 27  
RESULT 41  
RS17\_CHLPP  
ID RS17\_CHLPP STANDARD; PRT; 86 AA.  
AC Q9Z7R6; Q9JQG6;  
DT 30-MAY-2000 (Rel. 39, Created)  
DT 30-MAY-2000 (Rel. 39, Last sequence update)





DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Hypothetical protein MJ0492.  
 GN MJ0492.  
 OS Methanococcus jannaschii.  
 OC Archaea: Euryarchaeota; Methanococci; Methanococcales;  
 OC Methanocaldococcaceae; Methanocaldococcus.  
 OX NCBI\_TaxID=2190;  
 RN [1]  
 RN SEQUENCE FROM N.A.  
 RP STRAIN=JAL-1 / DSM 2661 / ATCC 43067;  
 RC MEDLINE=96337999; PubMed=8688087;  
 RA Bult C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,  
 RA Sutton G.G., Blake J.A., FitzGerald L.M., Clayton R.A., Gocayne J.D.,  
 RA Kierlavage A.R., Dougherty B.A., Tomb J.-F., Adams M.D., Reich C.I.,  
 RA Overbeek R., Kirkness E.F., Weinstock K.G., Merrick J.M., Glodek A.,  
 RA Scott J.L., Geoghegan N.S.M., Weidman J.F., Fuhrmann J.L., Nguyen D.,  
 RA Uitterback T.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,  
 RA Cotton M.D., Roberts K.M., Hurst M.A., Kaine B.P., Borodovsky M.,  
 RA Klenk H.-P., Fraser C.M., Smith H.O., Woese C.R., Venter J.C.;  
 RA "Complete genome sequence of the methanogenic archaeon, Methanococcus  
 RT jannaschii.";  
 RT Science 273:1058-1073 (1996).  
 RL -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL; U67499; AAB98488.1; -.  
 DR PIR; D64361; D64361.  
 DR TIGR; MJ0492; -.  
 DR HYPOHETHELICAL PROTEIN; Transmembrane; Complete proteome.  
 KW TRANSMEM 3 23 POTENTIAL.  
 FT SEQUENCE 95 AA; 10694 MW; 7A49FD4F888A365C CRC64;  
 SQ  
 Query Match 2.5%; Score 6; DB 1; Length 95;  
 Best Local Similarity 100.0%; Pred. No. 1.1e-02;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 158 VIIIAI 163  
 Db |||||  
 4 VIIIAI 9  
 RESULT 44  
 ID THIO ALIAC STANDARD; PRT; 105 AA.  
 AC P80579;  
 DT 01-OCT-1996 (Rel. 34, Created)  
 DT 01-OCT-1996 (Rel. 34, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Thioresoxin (TRX).  
 DE Thioresoxin (TRX).  
 GN TRXA.  
 OS Alicyclobacillus acidocaldarius (Bacillus acidocaldarius).  
 OC Bacteria; Firmicutes; Bacillales; Alicyclobacillaceae;  
 OC Alicyclobacillus.  
 OX NCBI\_TaxID=1388;  
 RN [1]  
 RN SEQUENCE, AND CHARACTERIZATION.  
 RP MEDLINE=98028685; PubMed=9359865;  
 RA Bartolucci S., Guagliardi A., Pedone E., de Pascale D., Cannio R.,  
 RA Canardella L., Rossi M., Nicastro G., de Chiara C., Facci P.,  
 RA Mascetti G., Nicolini C.;  
 RA "Thioresoxin from Bacillus acidocaldarius: characterization,  
 RT high-level expression in Escherichia coli and molecular modelling.";  
 RL Biochem. J. 328:277-285 (1997).  
 RN [2]  
 RN STRUCTURE BY NMR.  
 RP MEDLINE=20098516; PubMed=10632710;

RA Nicastro G., De Chiara C., Pedone E., Tato M., Rossi M.,  
 RA Bartolucci S.;  
 RA "NMR solution structure of a novel thioresoxin from bacillus  
 RT acidocaldarius possible determinants of protein stability.";  
 RL Eur. J. Biochem. 267:403-413 (2000).  
 CC -1- FUNCTION: Participates in various redox reactions through the  
 CC reversible oxidation of its active center dithiol to a disulfide  
 CC and catalyzes dithiol-disulfide exchange reactions.  
 CC -1- MASS SPECTROMETRY: MW=11577; METHOD=Electrospray.  
 CC -1- SIMILARITY: Belongs to the thioresoxin family.  
 DR PDB; 1QW; 26-JAN-00.  
 DR InterPro; IPR006662; Thioresox.  
 DR InterPro; IPR006663; Thioresox dom2.  
 DR InterPro; IPR005746; Thioresoxin.  
 DR Pfam; PF00085; Thioresox; 1.  
 DR PRINTS; PR00421; THIOREDOXIN.  
 DR TIGRFAMs; TIGR01068; thioresoxin; 1.  
 DR PROSITE; PS00194; THIOREDOXIN; 1.  
 DR Redox-active center; Electron transport; 3D-structure.  
 KW DISULFID 29 32 REDOX-ACTIVE.  
 FT HELIX 7 14  
 FT TURN 15 17  
 FT STRAND 20 25  
 FT TURN 27 28  
 FT TURN 31 31  
 FT HELIX 32 45  
 FT TURN 46 46  
 FT STRAND 52 56  
 FT TURN 57 59  
 FT HELIX 62 67  
 FT TURN 68 68  
 FT STRAND 74 78  
 FT STRAND 83 88  
 FT HELIX 93 105  
 SQ SEQUENCE 105 AA; 11576 MW; E03P636DFB3C3745 CRC64;  
 Query Match 2.5%; Score 6; DB 1; Length 105;  
 Best Local Similarity 100.0%; Pred. No. 1.2e-02;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 81 PETTSQ 86  
 Db |||||  
 61 PETTSQ 66  
 RESULT 45  
 ID RLA3 YEAST STANDARD; PRT; 106 AA.  
 AC P10622;  
 DT 01-JUL-1989 (Rel. 11, Created)  
 DT 01-JUL-1989 (Rel. 11, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE 60S acidic ribosomal protein P1-beta (L44') (L12EIIIB).  
 GN RPL18 OR RPLA3 OR L12EIIIB OR RPL44P OR YDL130W.  
 OS Saccharomyces cerevisiae (Baker's yeast).  
 OS Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
 OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.  
 OX NCBI\_TaxID=4932;  
 RN [1]  
 RN SEQUENCE FROM N.A.  
 RP MEDLINE=88243786; PubMed=2837476;  
 RA Remacha M., Saenz-Robles M.T., Vilella M.D., Ballesta J.P.G.;  
 RT "Independent genes coding for three acidic proteins of the large  
 RT ribosomal subunit from Saccharomyces cerevisiae.";  
 RL J. Biol. Chem. 263:9094-9101 (1988).  
 RN [2]  
 RN SEQUENCE FROM N.A.  
 RP STRAIN=SR26-12C;  
 RX MEDLINE=90130289; PubMed=2404943;  
 RA Newton C.H., Shimm L.C., Yee J., Dennis P.P.;  
 RT "A family of genes encode the multiple forms of the Saccharomyces  
 RT cerevisiae ribosomal proteins equivalent to the Escherichia coli L12  
 RT protein and a single form of the L10-equivalent ribosomal protein.";

```

RL J. Bacteriol. 172:579-588 (1990).
RN [3]
RP SEQUENCE FROM N.A.
RA Rieger M., Mueller-Auer S., Brueckner M., Schaefer M., Wagner G.;
RL Submitted (JUL-1996) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: Plays an important role in the elongation step of
CC protein synthesis.
CC -!- SUBUNIT: P1 and P2 exist as dimers at the large ribosomal subunit.
CC -!- MISCELLANEOUS: Yeasts contain 4 individual small ribosomal A
CC proteins (RPA) which can be classified into two couples of similar
CC but not identical sequences. Each couple is distinctly related to
CC one of the two A proteins present in multicellular organisms.
CC -!- SIMILARITY: Belongs to the L12P family of ribosomal proteins.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; M19238; AAA34973.1; -.
CC EMBL; M26507; AAA34734.1; -.
CC EMBL; Z74178; CAA98698.1; -.
CC PIR; C28104; R8BY28.
CC GerMOnline; 140372; -.
CC SGD; S0002288; RPPIB.
CC InterPro; IPR001813; Ribosomal_60S.
CC Pfam; PF00428; 60S_Ribosomal; 1.
CC Ribosomal protein; Phosphorylation; Multigene family.
KW SEQUENCE 106 AA; 10667 MW; EAED4F748653E0DC CRC64;
SQ
Query Match 2.5%; Score 6; DB 1; Length 106;
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 176 GKLIKE 181
Db |||||
50 GKOLKE 55
RESULT 46
THCC_RHOER STANDARD; PRT; 106 AA.
AC P43493;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Rhodocoxin.
DE THCC.
OS Rhodococcus erythropolis.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Nocardiaceae; Rhodococcus.
OX NCBI_TaxID=1833;
[1]
RN SEQUENCE FROM N.A., AND SEQUENCE OF 1-7.
RC STRAIN-NH6/21;
RA MEDLINE=95138028; PubMed=7836301;
RA Nagy I., Schoofs G., Compennolle F., Proost P., Vanderleyden J.,
RA de Mot R.;
RA "Degradation of the thiocarbamate herbicide EPTC (S-ethyl
RT dipropylcarbamothioate) and biosafening by Rhodococcus sp. strain
RT NH6/21 involve an inducible cytochrome P-450 system and aldehyde
RT dehydrogenase."
RL J. Bacteriol. 177:676-687 (1995).
CC -!- FUNCTION: FERREDOXIN-TYPE PROTEIN WHICH TRANSFERS ELECTRONS FROM
CC RHODOCOXIN REDUCTASE TO CYTOCHROME CYP116 (THCB), WHICH IS
CC INVOLVED IN THE DEGRADATION OF THIOCARBAMATE HERBICIDES.
CC -!- COFACTOR: Binds 1 2Fe-2S cluster (By similarity).
CC -!- SIMILARITY: Belongs to the adrenodoxin / putidaredoxin family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration

```

```

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; U17130; AAC45751.1; -.
CC HSSP; P00259; IGPX.
CC InterPro; IPR001055; Adrenodoxin.
CC Pfam; PF001041; Ferrdoxin.
CC Pfam; PF00111; fer2; 1.
CC PRINTS; PR00355; ADRENODOXIN.
CC PROSITE; PS00814; ADX; 1.
KW Electron transport; Metal-binding; Iron-sulfur; Iron; 2Fe-2S.
FT INIT MET 0
FT METAL 39 39 IRON-SULFUR (2FE-2S) (BY SIMILARITY).
FT METAL 45 45 IRON-SULFUR (2FE-2S) (BY SIMILARITY).
FT METAL 48 48 IRON-SULFUR (2FE-2S) (BY SIMILARITY).
FT METAL 86 86 IRON-SULFUR (2FE-2S) (BY SIMILARITY).
SQ SEQUENCE 106 AA; 11415 MW; AE5AAB91CD5710B CRC64;
Query Match 2.5%; Score 6; DB 1; Length 106;
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 162 AIGAGI 167
Db |||||
27 AIGAGI 32
RESULT 47
DBH_BORAD STANDARD; PRT; 108 AA.
ID DBH_BORAD
AC Q44625;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE DNA-binding protein HBBU.
DE HBB.
OS Borrelia andersonii.
OC Bacteria; Spirochaetes; Spirochaetales; Spirochaetaceae; Borrelia.
OX NCBI_TaxID=42109;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-19952;
RA Valsangiacomo C., Balmelli T., Piffaretti J.C.;
RL Submitted (FEB-1996) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: Histone-like DNA-binding protein which is capable of
CC wrapping DNA to stabilize it, and thus to prevent its denaturation
CC under extreme environmental conditions (By similarity).
CC -!- SIMILARITY: Belongs to the bacterial histone-like protein family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; U48686; AAC73108.1; -.
CC HSSP; P02346; IHUE.
CC InterPro; IPR000119; Bac DNABind.
CC Pfam; PF00216; Bac DNA binding; 1.
CC PRINTS; PR01727; DNABINDINGHU.
CC ProDom; PD000945; Bac DNABind; 1.
CC SMART; SM00411; BHL; 1
CC PROSITE; PS00045; HISTONE LIKE; FALSE_NEG.
KW DNA-binding; DNA denaturation.
SQ SEQUENCE 108 AA; 12742 MW; E25A6F802B361B88 CRC64;
Query Match 2.5%; Score 6; DB 1; Length 108;

```

```

Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 176 GKDLKE 181
DB 96 GKDLKE 101

RESULT 48
DBH BORAF
ID DBH BORAF STANDARD; PRT; 108 AA.
AC Q5720; Q44624;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE DNA-binding protein HBBU.
GN HBB.
OS Borrelia afzelii.
OC Bacteria; Spirochaetes; Spirochaetales; Spirochaetaceae; Borrelia.
OX NCBI_TaxID=29518;
RN [1]
SEQUENCE FROM N.A.
STRAIN=A268, B023, DK8, ECM1, U01, and V8461;
RA Valsangiacomo C., Balmelli T., Piffaretti J.C.;
RL Submitted (FEB-1996) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: Histone-like DNA-binding protein which is capable of
CC wrapping DNA to stabilize it, and thus to prevent its denaturation
CC under extreme environmental conditions (By similarity).
CC -!- SIMILARITY: Belongs to the bacterial histone-like protein family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; U48671; AAC73093.1; -
CC EMBL; U48672; AAC73094.1; -
CC EMBL; U48673; AAC73095.1; -
CC EMBL; U48674; AAC73096.1; -
CC EMBL; U48675; AAC73097.1; -
CC EMBL; U48676; AAC73098.1; -
CC EMBL; U48677; AAC73099.1; -
CC HSP; P02346; IHU.
CC InterPro; IPR000119; Bac DNABind.
CC Pfam; PF00216; Bac DNA binding; 1.
CC PRINTS; PR01727; DNABINDINGHU.
CC ProDom; PD000945; Bac DNABind; 1.
CC SMART; SM00411; BHL; 1.
CC PROSITE; PS00045; HISTONE LIKE; FALSE_NEG.
CC DNA-binding; DNA condensation.
CC VARIANT 19 19 S -> A (IN STRAIN U01).
CC VARIANT 23 23 R -> K (IN STRAIN U01).
CC VARIANT 89 89 H -> N (IN STRAIN U01).
CC SEQUENCE 108 AA; 12724 MW; 8B96BE73E40AEB1 CRC64;

Query Match 2.5%; Score 6; DB 1; Length 108;
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 176 GKDLKE 181
DB 96 GKDLKE 101

RESULT 49
DBH BORAF
ID DBH BORAF STANDARD; PRT; 108 AA.
AC Q57267; Q44834; Q44835; Q44837; Q44838; Q44839; Q57030;
AC Q57056; Q57402;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)

```

```

DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE DNA-binding protein HU (HBBU).
GN HUP OR HBB OR B0232.
OS Borrelia burgdorferi (Lyme disease spirochete).
OC Bacteria; Spirochaetes; Spirochaetales; Spirochaetaceae; Borrelia.
OX NCBI_TaxID=139;
RN [1]
SEQUENCE FROM N.A.
MEDLINE=93197131; PubMed=8451174;
RA Tilly K., Campbell J.;
RT "A Borrelia burgdorferi homolog of the Escherichia coli rho gene.";
RL Nucleic Acids Res. 21:1040-1040(1993).
RN [2]
SEQUENCE FROM N.A.
STRAIN=Various strains;
RA Valsangiacomo C., Balmelli T., Piffaretti J.C.;
RL Submitted (FEB-1996) to the EMBL/GenBank/DBJ databases.
RN [3]
SEQUENCE FROM N.A.
STRAIN=ATCC 35210 / B31;
MEDLINE=98065943; PubMed=9403685;
RA Fraser C.M., Casjens S., Huang W.M., Sutton G.G., Clayton R.A.,
RA Lathigra R., White O., Ketchum K.A., Dodson R., Hickey E.K., Gwinn M.,
RA Dougherty B., Tomb J.-F., Fleischmann R.D., Richardson D.,
RA Peterson J., Karpavicz A.R., Quackenbush J., Salzberg S., Hanson M.,
RA van Vugt R., Palmer N., Adams M.D., Gocayne J.D., Weidman J.,
RA Uterback T., Matthey L., McDonald L., Artiach P., Bowman C.,
RA Garland S., Fujii C., Cotton M.D., Horst K., Roberts K., Hatch B.,
RA Smith H.O., Venter J.C.;
RT "Genomic sequence of a Lyme disease spirochaete, Borrelia
RT burgdorferi".
RL Nature 390:580-586(1997).
CC -!- FUNCTION: Histone-like DNA-binding protein which is capable of
CC wrapping DNA to stabilize it, and thus to prevent its denaturation
CC under extreme environmental conditions (By similarity).
CC -!- SIMILARITY: Belongs to the bacterial histone-like protein family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; U35673; AAB41461.1; -
CC EMBL; U48648; AAC73070.1; -
CC EMBL; U48649; AAC73071.1; -
CC EMBL; U48650; AAC73072.1; -
CC EMBL; U48651; AAC73073.1; -
CC EMBL; U48652; AAC73074.1; -
CC EMBL; U48653; AAC73075.1; -
CC EMBL; U48654; AAC73076.1; -
CC EMBL; U48655; AAC73077.1; -
CC EMBL; U48656; AAC73078.1; -
CC EMBL; U48657; AAC73079.1; -
CC EMBL; U48658; AAC73080.1; -
CC EMBL; U48659; AAC73081.1; -
CC EMBL; U48660; AAC73082.1; -
CC EMBL; U48661; AAC73083.1; -
CC EMBL; U48662; AAC73084.1; -
CC EMBL; U48663; AAC73085.1; -
CC EMBL; U48664; AAC73086.1; -
CC EMBL; U48665; AAC73087.1; -
CC EMBL; U48666; AAC73088.1; -
CC EMBL; U48667; AAC73089.1; -
CC EMBL; U48668; AAC73090.1; -
CC EMBL; U48669; AAC73091.1; -
CC EMBL; U48670; AAC73092.1; -
CC EMBL; U48671; AAC73093.1; -
CC EMBL; U48672; AAC73094.1; -
CC EMBL; U48673; AAC73095.1; -
CC EMBL; U48674; AAC73096.1; -
CC EMBL; U48675; AAC73097.1; -
CC EMBL; U48676; AAC73098.1; -
CC EMBL; U48677; AAC73099.1; -
CC TIGR; B0232; -
CC InterPro; IPR000119; Bac DNABind.
CC Pfam; PF00216; Bac DNA binding; 1.
CC PRINTS; PR01727; DNABINDINGHU.
CC ProDom; PD000945; Bac DNABind; 1.
CC PROSITE; PS00045; HISTONE LIKE; FALSE_NEG.
CC DNA-binding; DNA condensation; Complete proteome.
CC VARIANT 4 4 S -> P (IN STRAINS UK AND V5116).
CC VARIANT 9 9 V -> I (IN STRAINS UK AND V5116).
CC VARIANT 9 9

```

```

Best Local Similarity 100.0%; Pred. No. 1.2e+02; Gaps 0;
Matches 6; Conservative 0; Mismatches 0; Indels 0;

QY 176 GKDLKE 181
DB 96 GKDLKE 101
|||||
|||||

RESULT 51
DBH_BORJA STANDARD; PRT; 108 AA.
ID Q45231; Q45227; Q45228; Q45229; Q45230;
AC 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE DNA-binding protein HBB.
GN HBB.
OS Borrelia japonica.
OC Bacteria; Spirochaetes; Spirochaetales; Spirochaetaceae; Borrelia.
OX NCBI_TaxID=34095;
RN [1]
SEQUENCE FROM N.A.
RC STRAIN=0612, H014, F63B, COW611A, and COW611C;
RL Valsangtacono C., Balmelli T., Piffaretti J.C.;
RA Submitted (FEB-1996) to the EMBL/GenBank/DBJ databases.
CC -i- FUNCTION: Histone-like DNA-binding protein which is capable of
CC wrapping DNA to stabilize it, and thus to prevent its denaturation
CC under extreme environmental conditions (By similarity)
CC -i- SIMILARITY: Belongs to the bacterial histone-like protein family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation
CC at the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announcement
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; U48677; AAC73099.1; -
CC EMBL; U48678; AAC73100.1; -
CC EMBL; U48679; AAC73101.1; -
CC EMBL; U48680; AAC73102.1; -
CC EMBL; U48681; AAC73103.1; -
CC EMBL; U48682; AAC73103.1; -
CC HSP; P02346; IHU.
CC Interpro: IPR000119; Bac_DNABind.
CC Pfam: PF00216; Bac_DNA_binding; 1.
CC PRINTS; PR01727; DNABINDINGHU.
CC ProDom; PD000945; Bac_DNABIND.
CC SMART; SM00411; BHL; 1.
CC PROSITE; PS00045; HISTONE_LIKE; FALSE_NEG.
CC DNA-binding; DNA condensation.
KW DNA-binding; 108 AA; 12684 MW; 90BBA42802B220F9E CRC64;
SQ SEQUENCE 108 AA; 2.5%; Score 6; DB 1; Length 108;
Query Match 2.5%; Score 6; DB 1; Length 108;
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 176 GKDLKE 181
DB 96 GKDLKE 101
|||||
|||||

RESULT 52
DBH_BORTU STANDARD; PRT; 108 AA.
ID Q45721;
AC Q45722;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE DNA-binding protein HBB.
GN HBB.
OS Borrelia turicatae.
OC Bacteria; Spirochaetes; Spirochaetales; Spirochaetaceae; Borrelia.

```

```

OX NCBI_TaxID=142;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=QM2007;
RA Valsangiaco C., Balmelli T., Piffaretti J.C.;
RL Submitted (FEB-1996) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: Histone-like DNA-binding protein which is capable of
CC wrapping DNA to stabilize it, and thus to prevent its denaturation
CC under extreme environmental conditions (by similarity).
CC -!- SIMILARITY: Belongs to the bacterial histone-like protein family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U48682; AAC73104.1; -
DR HSP; P02345; LHUE
DR InterPro; IPR000119; Bac_DNABind
DR Pfam; PF00216; Bac_DNA_Binding; 1.
DR PRINTS; PR01727; DNABINDINGHU.
DR ProDom; PD000945; Bac_DNABind; 1.
DR SMART; SM00411; BHL; 1.
DR PROSITE; PS00045; HISTONE LIKE; FALSE_NEG.
KW DNA-binding; DNA condensation.
SQ SEQUENCE 108 AA; 12676 MW; 6B91C5E37A72BD59 CRC64;

Query Match 2.5%; Score 6; DB 1; Length 108;
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 176 GKDLKE 181
Db 96 GKDLKE 101

RESULT 53
RLAL SCHPO STANDARD; PRT; 109 AA.
AC P17476.
DT 01-AUG-1990 (Rel. 15, Created)
DT 01-AUG-1990 (Rel. 15, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE 60S acidic ribosomal protein P1-alpha 1 (Al).
GN RPA1 OR SPAC644.15.
OS Schizosaccharomyces pombe (Fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC Schizosaccharomycetes.
OX NCBI_TaxID=4896;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=90220620; PubMed=2325655;
RA Beltrame M., Bianchi M.B.;
RT "A gene family for acidic ribosomal proteins in Schizosaccharomycetes
RT pombe: two essential and two nonessential genes.";
RL Mol. Cell. Biol. 10:2341-2348(1990).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=94252568; PubMed=8194753;
RA Jiang Y.K., Jin Y.H., Kim E.M., Hong S.H., Fabre F., Park S.D.;
RT "Cloning and sequence analysis of rhp51+, a Schizosaccharomycetes pombe
RT homolog of the Saccharomyces cerevisiae RAD51 gene.";
RL Genes 142:207-211(1994).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=972;
RX MEDLINE=21848401; PubMed=11859360;
RA Wood V., Gwilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,
RA Sgouros J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,

```

```

RA Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,
RA Collins M., Connor R., Cronin A., Davis P., Feltwell T., Fraser A.,
RA Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,
RA Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagels K.,
RA James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,
RA Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odell C.,
RA Oliver K., O'Neill S., Pearson D., Quail M.A., Rabinowitsch E.,
RA Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,
RA Skelton J., Simmonds M., Squares R., Squares S., Stevens K.,
RA Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,
RA Woodward J., Volkart G., Kert R., Robben J., Grymonprez B.,
RA Weltjens I., Vanstreels E., Rieger M., Schaefer M., Mueller-Auer S.,
RA Gabel C., Fuchs M., Fritz C., Holzer E., Moestl D., Hilbert H.,
RA Bozzym K., Langer I., Beck A., Lehach H., Reinhardt R., Fohl T.M.,
RA Eger P., Zimmermann W., Wedler H., Wambutt R., Purnelle B.,
RA Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaure V., Mottier S.,
RA Galibert F., Ames S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,
RA Lucas M., Rochet M., Gaillardin C., Tallada V.A., Garzon A., Thode G.,
RA Daga R.R., Cruzado I., Jimenez J., Sanchez M., del Rey P., Benito J.,
RA Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Foreburg S.L.,
RA Cerrutti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,
RA Shpakovski G.V., Uesery D., Barrell B.G., Nurse P.;
RT "The genome sequence of Schizosaccharomycetes pombe.";
RL Nature 415:871-880(2002).
RN [4]
RP SEQUENCE OF 1-24 FROM N.A.
RX MEDLINE=94051565; PubMed=8233794;
RA Muris D.F.R., Vreeken K., Carr A.M., Broughton B.C., Lehmann A.R.,
RA Lohman P.H.M., Pastink A.;
RT "Cloning the RAD51 homologue of Schizosaccharomycetes pombe.";
RL Nucleic Acids Res. 21:4586-4591(1993).
CC -!- FUNCTION: Plays an important role in the elongation step of
CC protein synthesis.
CC -!- SUBUNIT: P1 and P2 exist as dimers at the large ribosomal subunit.
CC -!- MISCELLANEOUS: Yeasts contain 4 individual small ribosomal A
CC proteins (RPA) which can be classified into two couples of similar
CC but not identical sequences. Each couple is distinctly related to
CC one of the two A proteins present in multicellular organisms.
CC -!- MISCELLANEOUS: Rpa3 and rpa4 are essential for cell survival,
CC whereas rpa1 and rpa2 are not.
CC -!- SIMILARITY: Belongs to the L12P family of ribosomal proteins.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M33137; AAA35334.1; -
DR EMBL; Z24756; CAA80880.2; ALT_SEQ.
DR EMBL; AL355012; CAB90142.1; -
DR EMBL; Z22691; CAA80400.1; -
DR PIR; A34715; R6BV11
DR GeneDB; Spombe; SPAC644.15; -
DR GO; GO:0006412; P:protein biosynthesis; ISS.
DR InterPro; IPR001813; Ribosomal_60S.
DR InterPro; IPR001859; Ribosomal_P2.
DR Pfam; PF00428; 60S ribosomal; 1.
DR PRINTS; PR00456; RIBOSOMALP2.
KW Ribosomal protein; Phosphorylation; Multigene family.
FT CONFLICT 24 24 S -> V (IN REF. 4).
SQ SEQUENCE 109 AA; 11141 MW; 2C5878183226AA04 CRC64;

Query Match 2.5%; Score 6; DB 1; Length 109;
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 176 GKDLKE 181
Db 52 GKDLKE 57

```

```
RESULT 54
RLAS5_SCHPO STANDARD; PRT; 109 AA.
ID Q9UW78; O14316;
AC RLA5_SCHPO
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE 60S acidic ribosomal protein P1-alpha 5.
GN RPA5 OR RPPI-3 OR RPPI-5 OR SPC1E11.09C.
OS Schizosaccharomyces pombe (Fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC Schizosaccharomycetes.
OX NCBI_TaxID=4896;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=972;
RX MEDLINE=20165959; PubMed=10701132;
RT "Identification of rpa1-5 and rpa2-6 genes encoding two additional
RT variants of the 60S acidic ribosomal proteins of Schizosaccharomycetes
RT pombe."
RL Genome 43:205-207(2000).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=972;
RX MEDLINE=21848401; PubMed=11859360;
RA Wood V., Gwilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,
RA Scouras J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,
RA Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,
RA Collins M., Connor R., Cronin A., Davis P., Feltham D., Fraser A.,
RA Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,
RA Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagels K.,
RA James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,
RA Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odell C.,
RA Oliver K., O'Neill S., Pearson D., Quail M.A., Rabinowitsch E.,
RA Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,
RA Skelton J., Simmonds M., Squares R., Squares S., Stevens K.,
RA Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,
RA Woodward J., Volkert G., Aert R., Robben J., Grynoprez B.,
RA Weltjens I., Vansteenkiste E., Rieger M., Schaefer M., Mueller-Auer S.,
RA Gabel C., Fuchs M., Fritz C., Holzer E., Moestl D., Hilbert H.,
RA Borzym K., Langer I., Beck A., Leirach H., Reinhardt R., Pohl T.M.,
RA Eger P., Zimmermann W., Wedler H., Wambutt R., Purnelle B.,
RA Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaure V., Mottier S.,
RA Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,
RA Lucas M., Rochet M., Gaillardin C., Tallada V.A., Garzon A., Thode G.,
RA Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,
RA Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Foreburg S.L.,
RA Cerutti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,
RA Shpakovski G.V., Ussery D., Bartell B.G., Nurse P.;
RT "The genome sequence of Schizosaccharomycetes pombe."
RL Nature 415:871-880(2002).
CC -!- FUNCTION: Plays an important role in the elongation step of
CC protein synthesis (By similarity).
CC -!- SUBUNIT: P1 and P2 exist as dimers at the large ribosomal
CC subunit.
CC -!- SIMILARITY: Belongs to the L12P family of ribosomal proteins.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
```

```
DR PROSITE; PS00056; RIBOSOMAL_S17; 1.
KW Ribosomal protein; rRNA-binding; Complete proteome.
SQ SEQUENCE 109 AA; 11973 MW; 589B980365610A3C CRC64;

Query Match 2.5%; Score 6; DB 1; Length 109;
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 212 EKTUVV 217
DB 42 EKTUVV 47

RESULT 56
RLA3 SCHPO
ID -RLA3 SCHPO STANDARD; PRT; 110 AA.
DT 01-AUG-1990 (Rel. 15, Created)
DT 01-AUG-1990 (Rel. 15, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE 60S acidic ribosomal protein P1-alpha 3 (A3).
GN RPA3 OR SPBC3B9.13C.
OS Schizosaccharomyces pombe (Fission Yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC Schizosaccharomycetes.
OX NCBI_TaxID=4896;
[1]
SEQUENCE FROM N.A.
RX MEDLINE=21849401; PubMed=11859360;
RA Wood V., Williams R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,
RA Sgouras J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,
RA Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,
RA Collins M., Connor R., Cronin A., Davis P., Feltwell T., Fraser A.,
RA Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,
RA Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagels K.,
RA James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,
RA Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odell C.,
RA Oliver K., O'Neill S., Pearson D., Quail M.A., Rabinowitsch E.,
RA Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,
RA Skelton J., Simmonds M., Squares R., Squares S., Stevens K.,
RA Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,
RA Woodward J., Volckaert G., Aert R., Robben J., Grymonprez B.,
RA Welljens I., Vanstreels E., Rieger M., Schaefer M., Mueller-Auer S.,
RA Gabel C., Fuchs M., Fritz C., Holzer E., Moestl D., Hilbert H.,
RA Borzym K., Langer I., Beck A., Lehrach H., Reinhardt R., Pohl T.M.,
RA Eger P., Zimmermann W., Wedler H., Wambutt R., Purnelle B.,
RA Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaure V., Mottier S.,
RA Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,
RA Lucas M., Rochet M., Gaillardin C., Tallada V.A., Garzon A., Thode G.,
RA Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,
RA Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,
RA Cerrutini L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,
RA Shpakovski G.V., Ussery D., Barrell B.G., Nurse P.;
RL Nature 415:871-880(2002).
CC -!- FUNCTION: Plays an important role in the elongation step of
CC protein synthesis.
CC -!- SUBUNIT: P1 and P2 exist as dimers at the large ribosomal subunit.
CC -!- MISCELLANEOUS: Yeasts contain 4 individual small ribosomal A
CC proteins (RPA) which can be classified into two couples of similar
CC but not identical sequences. Each couple is distinctly related to
CC one of the two A proteins present in multicellular organisms.
CC -!- MISCELLANEOUS: Rpa3 and rpa4 are essential for cell survival,
CC whereas rpa1 and rpa2 are not.
```

```
CC -!- SIMILARITY: Belongs to the L12P family of ribosomal proteins.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; M33139; AAA35336.1; -.
CC EMBL; AL022070; CAA17793.1; -.
CC PIR; C34715; R6BYP3.
CC GeneDB; Spombe; SPBC3B9.13C; -.
CC InterPro; IPR001813; Ribosomal_60S.
CC Pfam; PF00428; 60S_ribosomal; 1.
CC Ribosomal protein; Phosphorylation; Multigene family.
CC SEQUENCE 110 AA; 11171 MW; 1734AC9779F5A891 CRC64;

Query Match 2.5%; Score 6; DB 1; Length 110;
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 176 GKDLKE 181
DB 52 GKDLKE 57

RESULT 57
RS17 HALMA
ID RS17 HALMA STANDARD; PRT; 111 AA.
AC P12741;
DT 01-OCT-1989 (Rel. 12, Created)
DT 01-AUG-1991 (Rel. 19, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE 30S ribosomal protein S17 (HmsA17) (HS14).
GN RPS17P.
OS Haloarcula marismortui (Halobacterium marismortui).
OC Archaea; Euryarchaeota; Halobacteria; Halobacteriales;
OC Halobacteriaceae; Haloarcula.
OX NCBI_TaxID=2238;
[1]
SEQUENCE FROM N.A.
RX MEDLINE=90336772; PubMed=2143141;
RA Arndt E.;
RA "Nucleotide sequence of four genes encoding ribosomal proteins from
RT the 'S10 and spectinomycin' operon equivalent region in the
RT archaeobacterium Halobacterium marismortui.";
RT FEBS Lett. 267:193-198(1990).
[2]
SEQUENCE.
RX MEDLINE=87309217; PubMed=3305503;
RA Kimura J., Kimura M.;
RA "The primary structures of ribosomal proteins S14 and S16 from the
RT archaeobacterium Halobacterium marismortui. Comparison with
RT eubacterial and eukaryotic ribosomal proteins.";
RA J. Biol. Chem. 263:12150-12157(1987).
CC -!- SIMILARITY: Belongs to the S17P family of ribosomal proteins.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; X55311; CAA39017.1; -.
CC PIR; S10733; R3HS17.
CC HSPSP; P23828; 1RIP.
CC InterPro; IPR008994; Nucleic acid OB.
CC InterPro; IPR000266; Ribosomal_S17.
CC Pfam; PF00366; Ribosomal_S17; 1.
```

```

DR PRINTS; PRO0973; RIBOSOMALS17.
DR ProDom; PD001295; Ribosomal_S17; 1.
DR PROSITE; PS00056; RIBOSOMAL_S17; 1.
KW Ribosomal protein; rRNA-binding.
FT INIT MET 0
FT CONFLICT 88 C -> S (IN REF. 2).
FT CONFLICT 92 S -> P (IN REF. 2).
FT CONFLICT 109 110 MISSING (IN REF. 2).
FT CONFLICT 111 AA; 12141 MW; C22DD05891C383A1 CRC64;
SQ SEQUENCE 111 AA; 12141 MW; C22DD05891C383A1 CRC64;

Query Match 2.5%; Score 6; DB 1; Length 111;
Best Local Similarity 100.0%; Pred. No. 1.2e+02; Indels 0; Gaps 0;
Matches 6; Conservative 0; Mismatches 0;

QY 212 EKTUVV 217
DB 41 EKTUVV 46

RESULT 58
IAA4_SORBI STANDARD; PRT; 118 AA.
ID IAA4_SORBI
AC P81367;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE Alpha-amylase inhibitor 4 (SI alpha-4).
OS Sorghum bicolor (Sorghum) (Sorghum vulgare).
ES Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC PACCAD clade; Panicoideae; Andropogoneae; Sorghum.
OC NCBI_TaxID=4558;
RX STRAIN=cv. French red; TISSUE=Seed;
RM MEDLINE=9314996; PubMed=1492093;
RA Bloch C. Jr.; Richardson M.;
RT "The amino acid sequences of two 13-kDa alpha-amylase inhibitors from
the seeds of Sorghum bicolor (L.) Moench."
RL Protein Seq. Data Anal. 5:27-30(1992).
CC -!- FUNCTION: Alpha-amylase inhibitor.
CC -!- SIMILARITY: Belongs to the cereal trypsin/alpha-amylase inhibitor
family.
CC HSSP; P01088; 1BEA.
DR InterPro; IPR003612; AAI.
DR InterPro; IPR006106; Amylase inhib.
DR InterPro; IPR006105; Iry/amy inhib.
DR Pfam; PF00234; tryp_alpha_aml; 1.
DR PRINTS; PRO0808; AMLASEINHTR.
DR SMART; SM00499; AAI; 1.
DR PROSITE; PS00426; CEREAL_TRYP_AMYL_INH; 1.
KW Alpha-amylase inhibitor.
FT DISULFID 7 60 BY SIMILARITY.
FT DISULFID 21 49 BY SIMILARITY.
FT DISULFID 30 82 BY SIMILARITY.
FT DISULFID 50 101 BY SIMILARITY.
SQ SEQUENCE 118 AA; 12499 MW; B9571100591806BF CRC64;

Query Match 2.5%; Score 6; DB 1; Length 118;
Best Local Similarity 100.0%; Pred. No. 1.3e+02; Indels 0; Gaps 0;
Matches 6; Conservative 0; Mismatches 0;

QY 190 REMQRI 195
DB 84 REMQRI 89

RESULT 59
YHAH_ECOLI STANDARD; PRT; 121 AA.
ID YHAH_ECOLI
AC P42621;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical protein yhaH.
GN YHAH OR B3103 OR Z4457 OR ECS3985.
OS Escherichia coli, and
OS Escherichia coli O157:H7;
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
NCBI_TaxID=562, 83334;
RN NCBI_TaxID=562, 83334;
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / MG1655;
RX MEDLINE=97426617; PubMed=9278503;
RA Blattner F.R.; Plunkett G. III; Bloch C.A.; Perna N.T.; Burland V.;
RA Riley M.; Collado-Vides J.; Glasner J.D.; Rode C.K.; Mayhew G.F.;
RA Gregor J.; Davis N.W.; Kirkpatrick H.A.; Goeden M.A.; Rose D.J.;
RA Mau B.; Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12."
RL Science 277:1453-1474(1997).
RN NCBI_TaxID=562, 83334;
RP SEQUENCE FROM N.A.
RC STRAIN=O157:H7 / EDL933 / ATCC 700927;
RX MEDLINE=21074935; PubMed=11206551;
RA Perna N.T.; Plunkett G. III; Burland V.; Mau B.; Glasner J.D.;
RA Rose D.J.; Mayhew G.F.; Evans P.S.; Gregor J.; Kirkpatrick H.A.;
RA Posfai G.; Hackett J.; Klink S.; Boutin A.; Shao Y.; Miller L.;
RA Grotbeck E.J.; Davis N.W.; Lim A.; Dimalanta E.T.; Potamousis K.;
RA Apodaca J.; Anantharaman T.S.; Lin J.; Yen G.; Schwartz D.C.;
RA Welch R.A.; Blattner F.R.;
RT "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7;"
RL Nature 409:529-533(2001).
RN NCBI_TaxID=562, 83334;
RP SEQUENCE FROM N.A.
RC STRAIN=O157:H7 / RIMD 050952;
RX MEDLINE=21156231; PubMed=11258796;
RA Hayashi T.; Makino K.; Ohnishi M.; Kurokawa K.; Ishii K.; Yokoyama K.;
RA Han C.-G.; Ohtsubo E.; Nakayama K.; Murata T.; Tanaka M.; Tobe T.;
RA Iida T.; Takami H.; Honda T.; Sasakiwa C.; Ogasawara N.; Yasunaga T.;
RA Kihara S.; Shibata T.; Hattori M.; Shinagawa H.;
RT "Complete genome sequence of enterohaemorrhagic Escherichia coli
O157:H7 and genomic comparison with a laboratory strain K-12."
RL DNA Res. 8:11-22(2001).
CC -!- SURCELLULAR LOCATION: Integral membrane protein (potential).
CC -!- SIMILARITY: STRONG, TO E.COLI YHAH.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC EMBL; U18997; AAA57907.1; ALT_FRAME.
DR EMBL; AE000392; AAC76138.1; -.
DR EMBL; AE005540; AAG58236.1; -.
DR EMBL; AP002564; BAB37408.1; -.
DR EMBL; AP1127; A91127.
DR FIR; D65099; D65099.
DR FIR; H85971; H85971.
DR EcoGene; EG12747; yhaH.
DR InterPro; IPR008523; DUF805.
DR Pfam; PF05656; DUF805; 1.
KW Hypothetical protein; Transmembrane; Complete proteome.
FT TRANSMEM 24 44 POTENTIAL.
FT TRANSMEM 50 70 POTENTIAL.
FT TRANSMEM 81 101 POTENTIAL.
SQ SEQUENCE 121 AA; 14281 MW; 6D1D3283B025061C CRC64;

Query Match 2.5%; Score 6; DB 1; Length 121;
Best Local Similarity 100.0%; Pred. No. 1.3e+02; Indels 0; Gaps 0;
Matches 6; Conservative 0; Mismatches 0;

QY 236 QACTPG 241

```



```
Db 104 QAGTGC 109
|||||
RESULT 60
RL7 LIBAS STANDARD; PRT; 122 AA.
AC P36247;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE 50S ribosomal protein L7/L12.
GN RPL7.
OS Liberibacter asiaticus (Liberibacter asiaticum).
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC Rhizobiaceae; Candidatus Liberibacter.
OX NCBI_TaxID=34021;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=93153297; PubMed=7763375;
RA Vallechanoux S., Garnier M., Laigret F., Renaudin J., Bove J.M.;
RT "The genome of the non-cultured, bacterial-like organism associated
RT with citrus greening disease contains the nusG-rplKAL-rpoBC gene
RT cluster and the gene for a bacteriophage type DNA polymerase.";
RL Curr. Microbiol. 26:161-166(1993)
CC -1- FUNCTION: Seems to be the binding site for several of the factors
CC involved in protein synthesis and appears to be essential for
CC accurate translation (By similarity).
CC -1- SIMILARITY: Belongs to the L12P family of ribosomal proteins.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
CC EMBL; X83681; CAA58652.1; -
CC PIR; S60044; S60044.
CC InterPro; IPR008972; Cupredoxin.
CC Modulation; Signal.
CC NON_TER 1
CC SIGNAL <1 12 POTENTIAL.
CC CHAIN 13 124 EARLY NODULIN 5.
CC DOMAIN 13 ? PLASTOCYANIN-LIKE.
CC DOMAIN 77 96 PRO-RICH.
CC SEQUENCE 124 AA; 13770 MW; 01C3464948625471 CRC64;
DR HSWP; P02392; ICTF.
DR HAWAP; MF_00368; -; 1.
DR InterPro; IPR008932; Ribos_L12/7_olig.
DR InterPro; IPR00206; Ribosomal_L12.
DR Pfam; PF00542; Ribosomal_L12; 1.
DR ProDom; PD001326; Ribosomal_L12; 1.
DR TIGRFAMs; TIGR00855; L12; 1.
KW Ribosomal protein.
SQ SEQUENCE 122 AA; 12955 MW; FC30A0AAC8D2926E CRC64;
Query Match 2.5%; Score 6; DB 1; Length 122;
Best Local Similarity 100.0%; Pred.No.1.3e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 36 ASAPVS 41
Db 34 ASAPVS 39
|||||
RESULT 61
NOS_VICSA STANDARD; PRT; 124 AA.
AC Q41705;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE Early nodulin 5 precursor (N-5) (Fragment).
GN ENOD5.
OS Vicia sativa (Spring vetch) (Tare).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC eurosids I; Fabales; Fabaceae; Papilionoideae; Viciae; Vicia.
OX NCBI_TaxID=3908;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=cv. Nigra; TISSUE=Root nodules;
MEDLINE=96011756; PubMed=7548828;
Vijn I., Yang W.-C., Pallisgaard N., Oestergaard Jensen E.,
van Kammen A., Bisseling T.;
"VseNOD5, VseNOD12 and VseNOD40 expression during Rhizobium-induced
nodule formation on Vicia sativa roots.";
Plant Mol. Biol. 28:1111-1119(1995).
CC -1- FUNCTION: Involved in the infection process during the plant-
CC rhizobium interaction.
CC -1- DEVELOPMENTAL STAGE: Expressed during Rhizobium-induced nodule
CC formation. In 4-day old nodules it is found in a small cluster of
CC cells in the primordium, and in this cluster infection threads are
CC present. At day 5, expression is seen in the complete central zone II
CC tissue. At day 20, expressed in the infected cells, and maximal
CC where it is only active in the proximal part of this zone. Levels
CC decrease to a lower level from one cell layer to another at the
CC transition of prefixation zone into interzone II-III and remains
CC at this reduced level in the fixation zone III.
CC -1- SIMILARITY: Contains 1 plastocyanin-like domain.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
CC EMBL; X83681; CAA58652.1; -
CC PIR; S60044; S60044.
CC InterPro; IPR008972; Cupredoxin.
CC Modulation; Signal.
CC NON_TER 1
CC SIGNAL <1 12 POTENTIAL.
CC CHAIN 13 124 EARLY NODULIN 5.
CC DOMAIN 13 ? PLASTOCYANIN-LIKE.
CC DOMAIN 77 96 PRO-RICH.
CC SEQUENCE 124 AA; 13770 MW; 01C3464948625471 CRC64;
Query Match 2.5%; Score 6; DB 1; Length 124;
Best Local Similarity 100.0%; Pred.No.1.3e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 195 IITPLS 200
Db 84 IITPLS 89
|||||
RESULT 62
CAL2_MOUSE STANDARD; PRT; 130 AA.
AC Q99MF3;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Calcitonin gene-related peptide II precursor (CGRP-II) (Beta-type
DE CGRP).
GN CALCB.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=429/SV;
MEDLINE=21604266; PubMed=11761712;
Thomas P.M., Nasonkin I., Zhang H., Gagel R.F., Cote G.J.;
RT "Structure of the mouse calcitonin/calcitonin gene-related peptide
RT alpha and beta genes.";
DNA Seq. 12:131-135(2001).
CC -1- FUNCTION: CGRP induces vasodilatation. It dilates a variety of
```

```
CC vessels including the coronary, cerebral and systemic vasculature.
CC Its abundance in the CNS also points toward a neurotransmitter or
CC neuromodulator role (By similarity).
CC -!- SUBCELLULAR LOCATION: Secreted.
CC -!- SIMILARITY: Belongs to the calcitonin family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: AF325526; AAK16431.1; -.
CC DR EMBL: AF325524; AAK16431.1; JOINED.
CC DR WGD: MGI:2151254; Calc.
CC DR InterPro: IPR001693; Calcitonin-like.
CC DR InterPro: IPR002163; Calcitonin-B.
CC DR Pfam: PF00214; Calc CGRP IAPP; 1.
CC DR PRINTS: PR00817; CALCITONIN.
CC DR SMART: SM00113; CALCITONIN; 1.
CC DR PROSITE: PS00258; CALCITONIN; 1.
CC DR Cleavage on pair of basic residues; Amidation; Hormone; Signal.
CC FT SIGNAL 1 26
CC FT PROPEP 27 82 BY SIMILARITY.
CC FT PEPTIDE 84 120 CALCITONIN GENE-RELATED PEPTIDE II.
CC FT PROPEP 127 130 BY SIMILARITY.
CC FT DISULFID 85 90 BY SIMILARITY.
CC FT MOD_RES 120 120 AMIDATION (G-121 PROVIDE AMIDE GROUP) (BY
CC FT SIMILARITY).
CC SEQUENCE 130 AA; 14623 MW; 97299244E8F6C536 CRC64;
CC -----
CC Query Match 2.5%; Score 6; DB 1; Length 130;
CC Best Local Similarity 100.0%; Pred. No. 1.4e+02;
CC Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
CC -----
CC 144 DLGLTG 149
CC 38 DLGLTG 43
CC -----
CC RESULT 63
CC RNPA_CORGL STANDARD; PRT; 133 AA.
CC AC QNLS1;
CC DT 28-FEB-2003 (Rel. 41, Created)
CC DT 28-FEB-2003 (Rel. 41, Last sequence update)
CC DT 28-FEB-2003 (Rel. 41, Last annotation update)
CC DE Ribonuclease P protein component (EC 3.1.26.5) (RNaseP protein) (RNase
CC P protein) (Protein C5).
CC GN RNPA OR CGL3098.
CC OS Corynebacterium glutamicum (Brevibacterium flavum).
CC OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
CC OC Corynebacterineae; Corynebacteriaceae; Corynebacterium.
CC OX NCBI_TaxID=1718;
CC RN [1]
CC RP SEQUENCE FROM N.A.
CC RC STRAIN=ATCC 13032 / DSM 20300 / NCIB 10025;
CC EA Nakagawa S.;
CC RT "Complete genomic sequence of Corynebacterium glutamicum ATCC 13032.";
CC RL Submitted (MAY-2002) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: RNaseP catalyzes the removal of the 5'-leader sequence
CC from pre-tRNA to produce the mature 5'terminus. It can also cleave
CC other RNA substrates such as 4.5S RNA. The protein component plays
CC an auxiliary but essential role in vivo by binding to the 5'-
CC leader sequence and broadening the substrate specificity of the
CC ribozyme (By similarity).
CC -!- CATALYTIC ACTIVITY: Endonucleolytic cleavage of RNA, removing 5'-
CC extra-nucleotide from tRNA precursor.
CC -!- SUBUNIT: Consists of a catalytic RNA component (M1 or rnpB) and a
CC protein subunit (By similarity).
CC -!- SIMILARITY: Belongs to the rnpA family.
```

```
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: AP005283; BAC00492.1; -.
CC DR HAMAP: MF 00227; -.
CC DR InterPro: IPR000100; Ribonuclease_P.
CC DR Pfam: PF00825; Ribonuclease_P; 1.
CC DR ProDom: PD003629; Ribonuclease_P; 1.
CC DR TIGRFAMs: TIGR00188; rnpA; 1.
CC DR PROSITE: PS00648; RIBONUCLEASE_P; FALSE NEG.
CC DR Hydrolase; Nuclease; Endonuclease; tRNA processing; RNA-binding;
CC KW Complete proteome.
CC SEQUENCE 133 AA; 14481 MW; 214D960133F8F4FB CRC64;
CC -----
CC Query Match 2.5%; Score 6; DB 1; Length 133;
CC Best Local Similarity 100.0%; Pred. No. 1.4e+02;
CC Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
CC -----
CC 213 KTVVVH 218
CC 27 KIVVVH 32
CC -----
CC RESULT 64
CC H32_TETPY STANDARD; PRT; 135 AA.
CC AC P15512;
CC DT 01-APR-1990 (Rel. 14, Created)
CC DT 01-APR-1990 (Rel. 14, Last sequence update)
CC DT 28-FEB-2003 (Rel. 41, Last annotation update)
CC DE Histone H3.2.
CC OS Tetrahymena pyriformis.
CC OC Eukaryota; Alveolata; Ciliophora; Oligohymenophorea; Hymenostomatida;
CC OC Tetrahymenina; Tetrahymena.
CC OX NCBI_TaxID=5908;
CC RN [1]
CC RP SEQUENCE.
CC RX MEDLINE=8428953; PubMed=6432775;
CC RA Hayashi T., Hayashi H., Fusauchi Y., Iwai K.;
CC RT "Tetrahymena histone H3. Purification and two variant sequences.";
CC RL J. Biochem. 95:1741-1749 (1984).
CC -!- FUNCTION: Histone H3, along with histone H4, plays a central role
CC in nucleosome formation.
CC -!- SUBUNIT: The nucleosome is an octamer containing two molecules
CC each of H2A, H2B, H3 and H4. The octamer wraps approximately 146
CC bp of DNA.
CC -!- SIMILARITY: Belongs to the histone H3 family.
CC PIR: B28852; B28852.
CC DR InterPro: IPR007124; Hist_TAF.
CC DR InterPro: IPR007125; Histone_core_D.
CC DR InterPro: IPR000164; Histone_H3.
CC DR Pfam: PF00125; histone; 1.
CC DR PRINTS: PR00622; HISTONEH3.
CC DR SMART: SM00428; H3; 1.
CC DR PROSITE: PS00322; HISTONE_H3_1; 1.
CC DR PROSITE: PS00359; HISTONE_H3_2; 1.
CC KW Nuclear protein; Chromosomal protein; DNA-binding; Nucleosome core;
CC MW Multigene family.
CC FT INIT MET 0
CC SEQUENCE 135 AA; 15388 MW; 45235F3F915595C CRC64;
CC -----
CC Query Match 2.5%; Score 6; DB 1; Length 135;
CC Best Local Similarity 100.0%; Pred. No. 1.4e+02;
CC Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
CC -----
CC 37 SAPVSG 42
CC 37 SAPVSG 42
CC -----
```

```

Db      28 SAPVSG 33

RESULT 65
H33 TETH
ID_ H33 TETH STANDARD; PRT; 135 AA.
AC P41353;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Histone H3.3 (H3.3)
GN HHT3.

OS Eukaryota; Alveolata; Ciliophora; Oligohymenophorea; Hymenostomatida;
OC Eukaryota; Alveolata; Ciliophora; Oligohymenophorea; Hymenostomatida;
OC Tetrahymena; Tetrahymena.
OC Tetrahymena; Tetrahymena.
OX NCBI_TaxID=5911;
RN [1]
RN SEQUENCE FROM N.A.
RX MEDLINE=94167244; PubMed=8121802;
RA Thacker T.H., Macgaffey J., Bowen J., Horowitz S., Shapiro D.L.,
RA Gorovsky M.A.;
RT "Independent evolutionary origin of histone H3.3-like variants of
RT animals and Tetrahymena"; 22:180-186(1994).
RL Nucleic Acids Res. 22:180-186(1994).
CC -1- FUNCTION: HISTONE H3, ALONG WITH HISTONE H4, PLAYS A CENTRAL ROLE
CC IN NUCLEOSOME FORMATION. THIS IS A MACRONUCLEAR REPLACEMENT
CC VARIANT.
CC -1- SUBUNIT: The nucleosome is an octamer containing two molecules
CC each of H2A, H2B, H3 and H4. The octamer wraps approximately 146
CC bp of DNA.
CC -1- SIMILARITY: Belongs to the histone H3 family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; M87305; AAC37188.1; -.
CC PIR; S41501; S41501.
CC InterPro; IPR007124; Hist.TAF.
CC InterPro; IPR007125; Histone core D.
CC InterPro; IPR000164; Histone_H3.
CC Pfam; PF00125; histone; 1.
CC PRINTS; PR00622; HISTONEH3.
CC SMART; SM00428; H3; 1.
CC PROSITE; PS00322; HISTONE_H3_1; 1.
CC PROSITE; PS00959; HISTONE_H3_2; 1.
CC Nuclear protein; Chromosomal protein; DNA-binding; Nucleosome core;
CC Multigene family.
CC INITIAT 0 BY SIMILARITY.
SQ SEQUENCE 135 AA; 15352 MW; 98235D1E7C155948 CRC64;

Query Match 2.5%; Score 6; DB 1; Length 135;
Best Local Similarity 100.0%; Pred. No. 1.4e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 37 SAPVSG 42
DB 28 SAPVSG 33

RESULT 66
ID_ N05_PEA
ID N05_PEA STANDARD; PRT; 135 AA.
AC P25226;
DT 01-MAY-1992 (Rel. 22, Created)
DT 01-MAY-1992 (Rel. 22, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE Early nodulin 5 precursor (N-5).
GN ENO5.

OS Pisum sativum (Garden pea).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC eucosids I; Fabales; Fabaceae; Papilionoideae; Viciaeae; Pisum.
OX NCBI_TaxID=3889;
RN [1]
RN SEQUENCE FROM N.A.
RX STRAIN=cv. Sparkle; TISSUE=Root nodules;
RX MEDLINE=93005665; PubMed=2152123;
RA Scheres B., van Engelen F., van der Knaap E., van de Wiel C.,
RA van Kammen A., Bisseling T.;
RT "Sequential induction of nodulin gene expression in the developing
RT pea nodule."; 2:687-700(1990).
RL Plant Cell 2:687-700(1990).
CC -1- FUNCTION: Involved in the infection process during the plant-
CC rhizobium interaction.
CC -1- TISSUE SPECIFICITY: Invasion zone and early symbiotic zone.
CC -1- DEVELOPMENTAL STAGE: Expressed in the second stage of root nodule
CC formation.
CC -1- SIMILARITY: Contains 1 plastocyanin-like domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; S45139; AAB23536.1; -.
CC PIR; JQ1084; JQ1084.
CC Nodulation; Signal.
FT SIGNAL 1 23
FT CHAIN 24 135
FT DOMAIN 24 135
FT DOMAIN 88 107
FT SEQUENCE 135 AA; 15063 MW; C6DC13D551FA32B CRC64;

Query Match 2.5%; Score 6; DB 1; Length 135;
Best Local Similarity 100.0%; Pred. No. 1.4e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 195 ITLPLS 200
DB 95 ITLPLS 100

RESULT 67
ID_ RUVX_DEIRA
ID RUVX_DEIRA STANDARD; PRT; 136 AA.
AC Q9RR12;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Putative Holliday junction resolvase (EC 3.1.1.-).
GN DR2509.

OS Deinococcus radiodurans.
OC Bacteria; Deinococcus-Thermus; Deinococci; Deinococcales;
OC Deinococcaceae; Deinococcus.
OX NCBI_TaxID=1299;
RN [1]
RN SEQUENCE FROM N.A.
RX STRAIN=RL / ATCC 13939 / DSM 20539 / NCIB 9279;
RX MEDLINE=20036896; PubMed=10567266;
RA White O., Eisen J.A., Heidelberg J.F., Hickey E.K., Peterson J.D.,
RA Dodson R.J., Haft D.H., Gwinn M.L., Nelson W.C., Richardson D.L.,
RA Moffat K.S., Qin H., Jiang L., Pamphile W., Crosby M., Shen M.,
RA Vamathevan J.J., Lam P., McDonald L., Utterback T., Zaleski C.,
RA Makarova K.S., Aravind L., Daly M.J., Minton K.W., Fleischmann R.D.,
RA Ketchum K.A., Nelson K.E., Salzberg S., Smith H.O., Venter J.C.,
RA Fraser C.M.;
RT "Genome sequence of the radioresistant bacterium Deinococcus
RT radiodurans RL.";

```

```
RL Science 286:1571-1577(1999).
CC -!- FUNCTION: Could be a nuclease that resolves Holliday junction
CC intermediates in genetic recombination.
CC -!- SUBCELLULAR LOCATION: Cytoplasmic (Potential).
CC -!- SIMILARITY: Belongs to the YGF HJR family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AE002080; AAF12050.1; -.
CC PIR; D75265; D75265.
CC TIGR; DR2509; -.
CC HAMAP; MF_00651; -.
CC InterPro; IPR005227; Cons_hypoth250.
CC InterPro; IPR006641; YGFC.
CC Pfam; PF03652; UPF0081; 1.
CC SMART; SMC0732; YGFC; 1.
CC TIGRFAMs; TIGR00250; TIGR00250; 1.
CC Hydrolyase; Nuclease; DNA repair; DNA recombination; Complete proteome.
CC SEQUENCE 136 AA; 14819 MW; 7595D5353D2F21DE CRC64;
CC -----
CC Query Match 2.5%; Score 6; DB 1; Length 136;
CC Best Local Similarity 100.0%; Pred. No. 1.5e+02;
CC Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
CC -----
QY 121 EAAAVQ 126
DB 123 EAAAVQ 128
DB -----
RESULT 68
RUVX HAEIN
ID RUVX_HAEIN STANDARD; PRT; 139 AA.
AC P43981;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Putative Holliday junction resolvase (EC 3.1.1.-).
GN HI0305.
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;
OC Pasteurellaceae; Haemophilus.
OX NCBI_TaxID=727;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Rd / KW20 / ATCC 51907;
RX MEDLINE=95350630; PubMed=7542800;
RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
RA Kervatage A.P., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,
RA McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,
RA Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M.,
RA Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,
RA Utterback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,
RA Fine L.D., Fritchman J.L., Fuhrmann J.L., Geoghagen N.S.M.,
RA Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,
RA Venter J.C.;
RA "Whole-genome random sequencing and assembly of Haemophilus influenzae
RA Rd.";
RL Science 269:496-512(1995).
CC -!- FUNCTION: Could be a nuclease that resolves Holliday junction
CC intermediates in genetic recombination.
CC -!- SUBCELLULAR LOCATION: Cytoplasmic (Potential).
CC -!- SIMILARITY: Belongs to the YGF HJR family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
```

```
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; U32716; AAC21970.1; -.
CC PIR; H64005; H64005.
CC TIGR; HI0305; -.
CC HAMAP; MF_00651; -.
CC InterPro; IPR005227; Cons_hypoth250.
CC InterPro; IPR006641; YGFC.
CC Pfam; PF03652; UPF0081; 1.
CC SMART; SMC0732; YGFC; 1.
CC TIGRFAMs; TIGR00250; TIGR00250; 1.
CC Hydrolyase; Nuclease; DNA repair; DNA recombination; Complete proteome.
CC SEQUENCE 139 AA; 15339 MW; A875112A732AC6B4 CRC64;
CC -----
CC Query Match 2.5%; Score 6; DB 1; Length 139;
CC Best Local Similarity 100.0%; Pred. No. 1.5e+02;
CC Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
CC -----
QY 86 QALPAF 91
DB 28 QALPAF 33
DB -----
RESULT 69
RUVX PASMU
ID RUVX_PASMU STANDARD; PRT; 140 AA.
AC Q9CUX0;
DT 10-OCT-2003 (Rel. 42, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Putative Holliday junction resolvase (EC 3.1.1.-).
GN PM1870.
OS Pasteurella multocida.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;
OC Pasteurellaceae; Pasteurella.
OX NCBI_TaxID=747;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Pm70;
RX MEDLINE=21145866; PubMed=11248100;
RA May B.J., Zhang Q., Li L.-I., Paustian M.L., Whittam T.S., Kapur V.;
RA "Complete genomic sequence of Pasteurella multocida Pm70.";
RL Proc. Natl. Acad. Sci. U.S.A. 98:3460-3465(2001).
CC -!- FUNCTION: Could be a nuclease that resolves Holliday junction
CC intermediates in genetic recombination.
CC -!- SUBCELLULAR LOCATION: Cytoplasmic (Potential).
CC -!- SIMILARITY: Belongs to the YGF HJR family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AF006224; AA03954.1; -.
CC HAMAP; MF_00651; -.
CC InterPro; IPR005227; Cons_hypoth250.
CC InterPro; IPR006641; YGFC.
CC Pfam; PF03652; UPF0081; 1.
CC SMART; SMC0732; YGFC; 1.
CC TIGRFAMs; TIGR00250; TIGR00250; 1.
CC Hydrolyase; Nuclease; DNA repair; DNA recombination; Complete proteome.
CC SEQUENCE 140 AA; 15254 MW; 62BE3F8B1B75EA07 CRC64;
CC -----
CC Query Match 2.5%; Score 6; DB 1; Length 140;
CC Best Local Similarity 100.0%; Pred. No. 1.5e+02;
CC Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
CC -----
QY 86 QALPAF 91
```

```

DB      28 QALPAF 33
|||||
RESULT 70
HBA2_TRICR STANDARD; PRT; 141 AA.
ID HBA2_TRICR
AC P10784;
DT 01-JUL-1989 (Rel. 11, Created)
DT 01-JUL-1989 (Rel. 11, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Hemoglobin alpha-2 chain (Minor)
OS Triturus cristatus (Great crested newt) (Warty newt)
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Caudata; Salamandridae; Salamandridae;
OC Triturus.
OC NCBI_TaxID=8323;
RN [1]
RP SEQUENCE
RX MEDLINE=89207117; PubMed=3242554;
RA Kleinschmidt T., Sgouros J.G., Braunitzer G.;
RT "The first sequenced normal hemoglobin lacking histidine in position
RT 146 of the beta-chains. The primary structures of the major and minor
RT hemoglobin components of the great crested newt (Triturus cristatus,
RT Urodela, Amphibia).";
RL Biol. Chem. Hoppe-Seyler 369:1343-1360(1988).
CC -!- FUNCTION: Involved in oxygen transport from the lung to the
CC various peripheral tissues.
CC -!- SUBUNIT: Minor hemoglobin is an heterotetramer of two alpha-2
CC chains and two beta-2 chains.
CC -!- TISSUE SPECIFICITY: Red blood cells.
CC -!- SIMILARITY: Belongs to the globin family.
PIR; S02026; S02026.
DR HSP; P01922; IZB0.
DR InterPro; IPR002338; Alpha haem.
DR InterPro; IPR000971; Globin.
DR Pfam; PF00042; Globin; 1.
DR PRINTS; PR00612; ALPHAHAE.
DR PROSITE; PS01033; GLOBIN; 1.
KW Heme; Oxygen transport; Transport; Erythrocyte.
FT METAL 58 58 IRON (HEME DISTAL LIGAND).
FT METAL 87 87 IRON (HEME PROXIMAL LIGAND).
SQ SEQUENCE 141 AA; 15714 MW; 4059AC571F483ED6 CRC64;

Query Match 2.5%; Score 6; DB 1; Length 141;
Best Local Similarity 100.0%; Pred. No. 1.5e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 176 GKDLKE 181
Db 45 GKDLKE 50
|||||

RESULT 71
RIBI_PHOLE STANDARD; PRT; 144 AA.
ID RIBI_PHOLE
AC Q01994;
DT 01-JUL-1993 (Rel. 26, Created)
DT 01-JUL-1993 (Rel. 26, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE 6,7-dimethyl-8-ribityllumazine synthase (EC 2.5.1.9) (DMRL synthase)
DE (Lumazine synthase) (Riboflavin synthase beta chain) (Fragment).
GN RIBH.
OS Photobacterium leiognathi.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales;
OC Vibrionaceae; Photobacterium.
OC NCBI_TaxID=658;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 25521;
RX MEDLINE=92360014; PubMed=1339274;
RA Lee C.Y., Meighen E.A.;
RT "The lux genes in Photobacterium leiognathi are closely linked with

genes corresponding in sequence to riboflavin synthesis genes.";
-!- FUNCTION: Riboflavin synthase is a bifunctional enzyme complex
catalyzing the formation of riboflavin from 5-amino-6-(1'-D)-
ribityl-amino-2,4(1H,3H)-pyrimidinone and L-3,4-dihydroxy-2-
butanone-4-phosphate via 6,7-dimethyl-8-lumazine. The beta subunit
catalyzes the condensation of 5-amino-6-(1'-D)-ribityl-amino-
2,4(1H,3H)-pyrimidinone with L-3,4-dihydroxy-2-butanone-4-
phosphate yielding 6,7-dimethyl-8-lumazine.
-!- CATALYTIC ACTIVITY: 2 6,7-dimethyl-8-(1'-D-ribityl)lumazine =
riboflavin + 4-(1'-D-ribitylamino)-5-amino-2,6-dihydroxypyrimidine.
-!- PATHWAY: Riboflavin biosynthesis; last step.
-!- SIMILARITY: Belongs to the DMRL synthase family.

This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
-----
EMBL; M90094; AAA73230.1; -.
PIR; P01110; P01110.
DR HSP; P11998; IRVW.
DR HAMAP; MF 00178; -.
DR InterPro; IPR002180; DMRL synthase.
DR Pfam; PF00885; DMRL synthase; 1.
DR ProDom; PD003664; DMRL synthase; 1.
DR TIGRFAMs; TIGR00114; RibH; 1.
KW Riboflavin biosynthesis; Transferase.
FT NON_TER 144 144
SQ SEQUENCE 144 AA; 15180 MW; 4A4E4820CD960359 CRC64;

Query Match 2.5%; Score 6; DB 1; Length 144;
Best Local Similarity 100.0%; Pred. No. 1.5e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 160 IIAIGA 165
Db 75 IIAIGA 80
|||||

RESULT 72
MBEL_ECOLI STANDARD; PRT; 148 AA.
ID MBEL_ECOLI
AC P08097;
DT 01-AUG-1988 (Rel. 08, Created)
DT 01-AUG-1988 (Rel. 08, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE Mobilization protein MOBB (Protein C).
GN MOBB OR C.
OS Escherichia coli.
OG Plasmid C10 DF13.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OC NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=87248075; PubMed=3596243;
RA van Putten A.J., Jochems G.J., de Lang R., Nijkamp H.J.J.;
RT "Structure and nucleotide sequence of the region encoding the
RT mobilization proteins of plasmid C10DF13."
RL Gene 51:171-178(1987).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=86314306; PubMed=3749334;
RA Nijkamp H.J.J., de Lang R., Stuitje A.R., van den Elsen P.J.M.,
RA Veltkamp E., van Putten A.J.;
RT "The complete nucleotide sequence of the bacteriocinogenic plasmid
RT C10DF13."
RL Plasmid 16:135-160(1986).
CC -!- FUNCTION: THIS PROTEIN IS ESSENTIAL TO PROMOTE THE SPECIFIC

```

```
CC TRANSFER OF THE PLASMID IN THE PRESENCE OF CONJUGATIVE PLASMIDS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; X04466; CAA28151.1; -.
CC PIR; B29050; MZEC6.
CC KW Mobility protein; Plasmid; Conjugation.
CC SQ SEQUENCE 148 AA; 15933 MW; 2AA49B9628586989 CRC64;
CC
CC Query Match 2.5%; Score 6; DB 1; Length 148;
CC Best Local Similarity 100.0%; Pred. No. 1.6e+02;
CC Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
CC
CC QY 123 AAVQPV 128
CC Db |||||
CC 52 AAVQPV 57
CC
CC RESULT 73
CC Y222 TREPA STANDARD; PRT; 148 AA.
CC AC Q83251;
CC DT 16-OCT-2001 (Rel. 40, Created)
CC DT 16-OCT-2001 (Rel. 40, Last sequence update)
CC DE Hypothetical protein TP0222 precursor.
CC GN TP0222.
CC OS Treponema pallidum.
CC OC Bacteria; Spirochaetes; Spirochaetales; Spirochaetaceae; Treponema.
CC OX NCBI_TaxID=160;
CC RN [1]
CC RP SEQUENCE FROM N.A.
CC RC STRAIN=Nichols;
CC RX MEDLINE=98332770; PubMed=9665876;
CC RA Fraser C.M., Norris S.J., Weinstock G.M., White O., Sutton G.G.,
CC Dadson R., Gwin M., Hickey E.K., Clayton R., Ketchum K.A.,
CC Sodergren E., Hardham J.M., McLeod M.P., Salzberg S., Peterson J.,
CC Khalak H., Richardson D., Howell J.K., Chidambaram M., Utterback T.,
CC McDonald L., Attiach P., Bowman C., Cotton M.D., Fujii C., Garland S.,
CC Hatch B., Horst K., Roberts K., Sandusky M., Weidman J., Smith H.O.,
CC Venter J.C.;
CC RA "Complete genome sequence of Treponema pallidum, the syphilis
CC spirochete";
CC RL Science 281:375-388(1998).
CC
CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AE001203; AAC65211.1; -.
CC PIR; C71353; C71353.
CC DR TIGR; TP0222; -.
CC KW Hypothetical protein; Transmembrane; Signal; Coiled coil;
CC Complete proteome.
CC FT SIGNAL 1 23 POTENTIAL.
CC FT CHAIN 24 148 HYPOTHETICAL PROTEIN TP0222.
CC FT DOMAIN 49 129 COILED COIL (POTENTIAL).
CC FT TRANSMEM 125 147 POTENTIAL.
CC SQ SEQUENCE 148 AA; 16224 MW; 389CE9EB3A5DA0BC CRC64;
CC
CC Query Match 2.5%; Score 6; DB 1; Length 148;
CC Best Local Similarity 100.0%; Pred. No. 1.6e+02;
CC Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
CC
```

QY 162 AICAGI 167  
Db |||||  
138 AICAGI 143

## RESULT 74

```
YUGK_ECOLI STANDARD; PRT; 150 AA.
ID YUGK_ECOLI
AC P39335;
DT 01-FEB-1995 (Rel. 31, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Protein YlgK.
GN YUGK OR R4252 OR SP4239 OR S4500.
OS Escherichia coli, and
OS Shigella flexneri.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=562, 623;
RN [1]
RP SEQUENCE FROM N.A.
RC SPECIES=E.coli; STRAIN=K12 / MG1655;
RX MEDLINE=95334362; PubMed=7610040;
RA Burland V.D., Plunkett G. III, Sofia H.J., Daniels D.L.,
RA Blattner F.R.;
RT "Analysis of the Escherichia coli genome VI: DNA sequence of the
RT region from 92.8 through 100 minutes.";
RL Nucleic Acids Res. 23:2105-2119(1995).
RN [2]
RP SEQUENCE FROM N.A.
RC SPECIES=S.flexneri; STRAIN=301 / Serotype 2a;
RX MEDLINE=22272406; PubMed=12384590;
RA Jin Q., Yuan Z., Xu J., Wang Y., Shen Y., Lu W., Wang J., Liu H.,
RA Sun J., Yang F., Zhang X., Zhang J., Yang G., Wu H., Qu D., Dong J.,
RA Sun L., Xue Y., Zhao A., Gao Y., Zhu J., Kan B., Ding K., Chen S.,
RA Cheng H., Yao Z., He B., Chen R., Ma D., Qiang B., Wen Y., Hou Y.,
RA Xu J.;
RT "Genome sequence of Shigella flexneri 2a: insights into pathogenicity
RT through comparison with genomes of Escherichia coli K12 and O157.";
RL Nucleic Acids Res. 30:4432-4441(2002).
RN [3]
RP SEQUENCE FROM N.A.
RC SPECIES=S.flexneri; STRAIN=2457T / ATCC 700930 / Serotype 2a;
RX MEDLINE=2590274; PubMed=12704152;
RA Wei J., Goldberg M.B., Burland V., Venkatesan M.M., Deng W.,
RA Fournier G., Mayhew G.F., Plunkett G. III, Rose D.J., Darling A.,
RA Mau B., Perna N.T., Payne S.M., Runyen-Janecky L.J., Zhou S.,
RA Schwartz D.C., Blattner F.R.;
RT "Complete genome sequence and comparative genomics of Shigella
RT flexneri serotype 2a strain 2457T.";
RL Infect. Immun. 71:2775-2786(2003).
RN [4]
RP IDENTIFICATION BY MASS SPECTROMETRY.
RC SPECIES=E.coli;
RX MEDLINE=99420866; PubMed=10493123;
RA Fountoulakis M., Takacs M.-F., Berndt P., Langen H., Takacs B.;
RT "Enrichment of low abundance proteins of Escherichia coli by
RT hydroxyapatite chromatography.";
RL Electrophoresis 20:2181-2195(1999).
CC -!- SIMILARITY: BELONGS TO THE YHCH/YTAL/YUGK FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; UL4003; AAA97148.1; ALT INIT.
CC EMBL; AE000496; AAC7209.1; ALT_INIT.
CC EMBL; AE015432; AAN45657.1; ALT_INIT.
```

```
DR EMBL; AB016993; AAP19444.1; -.
DR EcoGene; EGI2530; VJCK.
DR InterPro; IPR004375; Cons_hypoth22.
DR Pfam; PF04074; DUF386; 1.
DR TIGRFAMs; TIGR00022; TIGR00022; 1.
DR Complete proteome.
SQ SEQUENCE 150 AA; 16865 MW; 30E316B24C523DAE CRC64;

Query Match      2.5%; Score 6; DB 1; Length 150;
Best Local Similarity 100.0%; Pred. No. 1.6e-02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 210 VDEKTV 215
DB 107 VDEKTV 112

RESULT 75
SODC_NEUCR STANDARD; PRT; 153 AA.
AC P07509;
DT 01-APR-1988 (Rel. 07, Created)
DT 01-AUG-1991 (Rel. 19, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Superoxide dismutase [Cu-Zn] (EC 1.15.1.1).
GN SOD-1.
OS Neurospora crassa.
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
OC Sordariomycetidae; Sordariales; Sordariaceae; Neurospora.
OX NCBI_TaxID=5141;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=91035418; PubMed=2146266;
RA Chary P., Hallewell R.A., Natvig D.O.;
RT "Structure, exon pattern, and chromosome mapping of the gene for
RT cytosolic copper-zinc superoxide dismutase (sod-1) from Neurospora
RT crassa.";
RL J. Biol. Chem. 265:18961-18967(1990).
[2]
RP SEQUENCE.
RX MEDLINE=85261356; PubMed=3160699;
RA Lerch K., Schenk E.;
RT "Primary structure of copper-zinc superoxide dismutase from
RT Neurospora crassa.";
RL J. Biol. Chem. 260:9559-9566(1985).
CC -!- FUNCTION: Destroys radicals which are normally produced within the
CC cells and which are toxic to biological systems.
CC -!- CATALYTIC ACTIVITY: 2 superoxide + 2 H(+) = O(2) + H(2)O(2).
CC -!- COFACTOR: Binds 1 copper ion and 1 zinc ion per subunit (By
CC similarity).
CC -!- SUBUNIT: Homodimer.
CC -!- SUBCELLULAR LOCATION: Cytoplasmic.
CC -!- SIMILARITY: Belongs to the Cu-Zn superoxide dismutase family.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
DR EMBL; M58697; AA363780.1; -.
DR FIR; A36591; A36591.
DR HSP; P00445; IJCV.
DR InterPro; IPR001424; SOD_CU_ZN.
DR Pfam; PF00080; sodcu; 1.
DR PRINTS; PR00068; CUZNDISMTASE.
DR ProDom; PD000469; SOD_CU_ZN; 1.
DR PROSITE; PS00087; SOD_CU_ZN_1; 1.
DR PROSITE; PS00332; SOD_CU_ZN_2; 1.
KW Antioxidant; Oxidoreductase; Metal-binding; Copper; Zinc.
FT INIT_MET 0
```

```
FT METAL 46 46 COPPER (BY SIMILARITY).
FT METAL 48 48 COPPER (BY SIMILARITY).
FT METAL 63 63 COPPER AND ZINC (BY SIMILARITY).
FT METAL 71 71 ZINC (BY SIMILARITY).
FT METAL 80 80 ZINC (BY SIMILARITY).
FT METAL 83 83 ZINC (BY SIMILARITY).
FT METAL 120 120 COPPER (BY SIMILARITY).
FT DISULFID 57 146 BY SIMILARITY.
FT CONFLICT 69 69 T -> A (IN REF. 2).
SQ SEQUENCE 153 AA; 15753 MW; 56DAPAC86CD9573B CRC64;

Query Match      2.5%; Score 6; DB 1; Length 153;
Best Local Similarity 100.0%; Pred. No. 1.6e-02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 128 VIGISQ 133
DB 148 VIGISQ 153

Search completed: March 17, 2004, 07:25:19
Job time : 22 secs
```

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 17, 2004, 07:21:00 ; Search time 45 Seconds  
(without alignments)

1696.787 Million cell updates/sec

Title: US-10-057-951-2\_COPY\_22\_263

Perfect score: 242

Sequence: 1 SGGCFWDNGHLYREDTSPA.....PVDQEGSTPLMQAGTPGA 242

Scoring table: OLIGO

Gapop 60.0 , Gapext 60.0

Searched: 1017041 seqs, 315518202 residues

Word size : 0

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 100 summaries

Database :

SPTREMBL 25:\*

- 1: sp.archaea:\*
- 2: sp.bacteria:\*
- 3: sp.fungi:\*
- 4: sp.human:\*
- 5: sp.invertebrate:\*
- 6: sp.mammal:\*
- 7: sp.mhc:\*
- 8: sp.organelle:\*
- 9: sp.phase:\*
- 10: sp.plant:\*
- 11: sp.podent:\*
- 12: sp.virus:\*
- 13: sp.vertbrate:\*
- 14: sp.unclassified:\*
- 15: sp.rvirus:\*
- 16: sp.bacteriap:\*
- 17: sp.archaeap:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	242	100.0	263	4	Q96FE7 homo sapien
2	229	94.6	263	4	O00318 homo sapien
3	175	72.3	234	4	Q86YW2 homo sapien
4	101	41.7	263	4	Q8NCJ9 homo sapien
5	26	10.7	213	11	Q811Z3 mus musculus
6	26	10.7	264	11	Q811Z2 mus musculus
7	26	10.7	264	11	Q7TMJ8 mus musculus
8	11	4.5	263	13	Q7SXB3 brachydanio
9	9	3.7	331	16	Q99RQ5 Q8NV36 staphylococ
10	9	3.7	344	16	Q8NV36 Q800Y7 maleagris g
11	9	3.7	540	13	Q800Y7 Q805B4 bifidobacte
12	8	3.3	155	16	Q805B4 bifidobacte
13	8	3.3	345	8	Q85J73 varanus ros
14	8	3.3	452	16	Q815D6 bacillus an
15	8	3.3	452	16	Q815D6 bacillus ce
16	8	3.3	457	16	Q821K3 streptomyce

Q972d9 sulfolobus  
Q97eb2 clostridium  
P70521 rattus norv  
Q91xg8 mus musculu  
Q8pfy5 xanthomonas  
Q23587 caenorhabdi  
Q8wt10 plasmodium  
Q81k21 plasmodium  
Q8k63 methanosarc  
Q8pzm8 methanosarc  
Q13544 saccharomyc  
Q42428 latea calca  
Q94dn5 oryza sativ  
Q8tc13 methanosarc  
Q55181 synechocyst  
Q94xuz crotalus ve  
Q94xul crotalus du  
Q94xuz crotalus du  
Q8zk32 salmonella  
Q82124 salmonella  
Q8kue5 corynebacte  
Q8ntp9 corynebacte  
Q9c892 arabidopsis  
Q895v2 clostridium  
Q89cf0 bradyrhizob  
Q44013 alcaligenes  
Q98m02 rhizobium l  
Q931g1 propionibac  
Q9rt90 deinococcus  
P96683 bacillus su  
Q9ub2 schizosacch  
Q88it4 pseudomonas  
Q891q5 clostridium  
Q8bh49 mus musculu  
Q8xne0 clostridium  
Q8xne0 caulobacter  
Q9vqu4 drosophila  
Q9n3d7 caenorhabdi  
Q99wr0 staphylococ  
Q8nyc6 staphylococ  
Q834q7 enterococcu  
Q8pj49 xanthomonas  
Q8p8a4 xanthomonas  
Q986w0 rhizobium l  
Q97v08 sulfolobus  
Q9vvt5 drosophila  
Q23364 caenorhabdi  
Q89gi0 bradyrhizob  
Q17035 caenorhabdi  
Q8vq72 bacillus l1  
Q9j1b0 nelson bay  
Q8uik1 agrobacteri  
Q941w0 oryza sativ  
Q8yvt5 anabaena sp  
Q7vna0 haemophilus  
Q89pa5 bradyrhizob  
Q8yck8 anabaena sp  
Q8ky10 rhizobium e  
Q9gu11 caenorhabdi  
Q9j5a2 chlamydia p  
Q927b1 chlamydia p  
Q927a9 chlamydia p  
Q9j521 chlamydia p  
Q8pdm8 xanthomonas  
Q8d7w9 vitrio vuln  
Q82vf0 nitrosomona  
Q8h482 oryza sativ  
Q7vpu6 chlamydia p  
Q3br6 homo sapien  
Q8pkel xanthomonas  
Q8evd2 neisseria s  
Q9xc57 pseudomonas  
Q818w0 bacillus ce

17 17 Q972D9 475 3.3 8 3.3  
18 18 Q97EB2 602 3.3 8 3.3  
19 19 P70521 716 3.3 8 3.3  
20 20 Q91XG8 716 3.3 8 3.3  
21 21 Q8PFY5 784 3.3 8 3.3  
22 22 Q23587 2585 3.3 8 3.3  
23 23 Q8WT10 37 2.9 8 2.9  
24 24 Q81K21 107 2.9 8 2.9  
25 25 Q8K63 112 2.9 8 2.9  
26 26 Q8PZM8 113 2.9 8 2.9  
27 27 Q13544 120 2.9 8 2.9  
28 28 Q42428 129 2.9 8 2.9  
29 29 Q94DN5 135 2.9 8 2.9  
30 30 Q8TLC3 139 2.9 8 2.9  
31 31 Q55181 144 2.9 8 2.9  
32 32 Q94XU2 147 2.9 8 2.9  
33 33 Q94XU1 147 2.9 8 2.9  
34 34 Q94XU3 147 2.9 8 2.9  
35 35 Q82K32 150 2.9 8 2.9  
36 36 Q8Z124 150 2.9 8 2.9  
37 37 Q8KUE6 151 2.9 8 2.9  
38 38 Q8NTP9 154 2.9 8 2.9  
39 39 Q9C892 164 2.9 8 2.9  
40 40 Q895V2 176 2.9 8 2.9  
41 41 Q89CF0 185 2.9 8 2.9  
42 42 Q44013 192 2.9 8 2.9  
43 43 Q98M02 197 2.9 8 2.9  
44 44 Q931G1 209 2.9 8 2.9  
45 45 Q9RT90 222 2.9 8 2.9  
46 46 P96683 226 2.9 8 2.9  
47 47 Q9UUB2 229 2.9 8 2.9  
48 48 Q88IT4 233 2.9 8 2.9  
49 49 Q891Q5 254 2.9 8 2.9  
50 50 Q8BH49 266 2.9 8 2.9  
51 51 Q8XNE0 272 2.9 8 2.9  
52 52 Q9VQU4 273 2.9 8 2.9  
53 53 Q9N3D7 282 2.9 8 2.9  
54 54 Q99WR0 286 2.9 8 2.9  
55 55 Q8NYC6 286 2.9 8 2.9  
56 56 Q834Q7 286 2.9 8 2.9  
57 57 Q8PJ49 288 2.9 8 2.9  
58 58 Q8PF84 288 2.9 8 2.9  
59 59 Q8P8A4 290 2.9 8 2.9  
60 60 Q986W0 291 2.9 8 2.9  
61 61 Q97V08 291 2.9 8 2.9  
62 62 Q9VVT5 302 2.9 8 2.9  
63 63 Q23364 303 2.9 8 2.9  
64 64 Q89GJ0 313 2.9 8 2.9  
65 65 Q17035 314 2.9 8 2.9  
66 66 Q8VQ72 322 2.9 8 2.9  
67 67 Q9J1B0 323 2.9 8 2.9  
68 68 Q8UIK1 326 2.9 8 2.9  
69 69 Q941W0 332 2.9 8 2.9  
70 70 Q8YWT5 332 2.9 8 2.9  
71 71 Q7VNA0 336 2.9 8 2.9  
72 72 Q89PA5 338 2.9 8 2.9  
73 73 Q8YRX8 346 2.9 8 2.9  
74 74 Q8KY10 351 2.9 8 2.9  
75 75 Q9GU11 355 2.9 8 2.9  
76 76 Q9J5A2 356 2.9 8 2.9  
77 77 Q927B1 356 2.9 8 2.9  
78 78 Q927A9 365 2.9 8 2.9  
79 79 Q9S221 365 2.9 8 2.9  
80 80 Q8PDM8 365 2.9 8 2.9  
81 81 Q8D7W9 370 2.9 8 2.9  
82 82 Q82VF0 371 2.9 8 2.9  
83 83 Q8H482 379 2.9 8 2.9  
84 84 Q7VPU6 381 2.9 8 2.9  
85 85 Q3BR6 393 2.9 8 2.9  
86 86 Q8PKE1 394 2.9 8 2.9  
87 87 Q3EVD2 396 2.9 8 2.9  
88 88 Q9XC57 398 2.9 8 2.9  
89 89 Q818W0 408 2.9 8 2.9



90 7 2.9 414 16 Q8CMV9  
 91 7 2.9 421 16 Q8DF75  
 92 7 2.9 426 10 Q8LQW3  
 93 7 2.9 430 16 Q98N05  
 94 7 2.9 448 10 Q9STM4  
 95 7 2.9 448 16 Q8UFR7  
 96 7 2.9 449 16 Q92QK1  
 97 7 2.9 451 4 Q8N6V2  
 98 7 2.9 451 16 Q8R7E2  
 99 7 2.9 453 17 Q8TQ98  
 100 7 2.9 454 16 Q9PK13

## ALIGNMENTS

RESULT 1  
 Q96FE7 PRELIMINARY; PRT; 263 AA.  
 AC Q96FE7;  
 DT 01-DEC-2001 (T-EMBLrel. 19, Created)  
 DT 01-DEC-2001 (T-EMBLrel. 19, Last sequence update)  
 DT 01-JUN-2003 (T-EMBLrel. 24, Last annotation update)  
 DE Hypothetical protein (HGFL(L) protein).  
 GN HGFL  
 OS Homo sapiens (Human)  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC Tissue=Brain;  
 RA Strausberg R.;  
 RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC Chiang H., Chang M.;  
 RA Submitted (JUL-2002) to the EMBL/GenBank/DBJ databases.  
 RL -1- SIMILARITY: CONTAINS 1 KRINGLE DOMAIN.  
 CC EMBL; BC011049; AAH1049.1; -.  
 DR EMBL; AF528880; AA033763.1; -.  
 DR InterPro; IPR000001; Kringle.  
 DR Pfam; PF00051; kringle; 1.  
 DR PRINTS; PR00018; KRINGLE.  
 DR ProDom; PD000395; Kringle; 1.  
 DR SMART; SMO0130; KR; 1.  
 DR PROSITE; PS00021; KRINGLE 1; 1.  
 DR PROSITE; PS50070; KRINGLE\_2; 1.  
 KW Hypothetical protein; Glycoprotein; Kringle.  
 SQ SEQUENCE 263 AA; 28234 MW; 197C3EEB888FA242 CRC64;  
 Query Match 100.0%; Score 242; DB 4; Length 263;  
 Best Local Similarity 100.0%; Pred. No. 1.2e-248;  
 Matches 242; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 SGGCFWNGHLYREDQTSAPAGRLCLNWLDAQSGLASAPVSGAGNSHCNPNDEPRGPW 60  
 DB 22 SGGCFWNGHLYREDQTSAPAGRLCLNWLDAQSGLASAPVSGAGNSHCNPNDEPRGPW 81  
 QY 61 CYVSGEAGVPEKPCEDLRCPEITTSQALPAFTTEIQEASEGPGADEVQVFAPANALPARS 120  
 DB 82 CYVSGEAGVPEKPCEDLRCPEITTSQALPAFTTEIQEASEGPGADEVQVFAPANALPARS 141  
 QY 121 EAAAVQPVIGISQVRVNSKEKDLGLTGLVGLITMMVIIIAIGAGIILGYSKRGKDLK 180  
 DB 142 EAAAVQPVIGISQVRVNSKEKDLGLTGLVGLITMMVIIIAIGAGIILGYSKRGKDLK 201  
 QY 181 EQHDKVCEREMORITLPLSAFTNPTCEIVDEKTVVHTSQTVPDQEGSTPLMGAGT 240  
 DB 202 EQHDKVCEREMORITLPLSAFTNPTCEIVDEKTVVHTSQTVPDQEGSTPLMGAGT 261  
 QY 241 GA 242

Db 262 GA 263  
 RESULT 2  
 Q00318 PRELIMINARY; PRT; 263 AA.  
 ID Q00318;  
 AC Q00318;  
 DT 01-JUL-1997 (T-EMBLrel. 04, Created)  
 DT 01-JUL-1997 (T-EMBLrel. 04, Last sequence update)  
 DT 01-OCT-2003 (T-EMBLrel. 25, Last annotation update)  
 DE WUGSC:DJ515N1.2 protein.  
 GN WUGSC:DJ515N1.2  
 OS Homo sapiens (Human)  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Du Z., Scheet P., Harper M.;  
 RT "The sequence of H. sapiens PAC clone RP3-515N1";  
 RL Submitted (MAY-1997) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RA Waterston R.;  
 RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.  
 CC -1- SIMILARITY: CONTAINS 1 KRINGLE DOMAIN.  
 DR EMBL; AC002073; AAB54054.1; -.  
 DR HSPF; P00749; IKDU.  
 DR InterPro; IPR000001; Kringle.  
 DR Pfam; PF00051; kringle; 1.  
 DR PRINTS; PR00018; KRINGLE.  
 DR ProDom; PD000395; Kringle; 1.  
 DR SMART; SMO0130; KR; 1.  
 DR PROSITE; PS00021; KRINGLE 1; 1.  
 DR PROSITE; PS50070; KRINGLE\_2; 1.  
 KW Glycoprotein; Kringle.  
 SQ SEQUENCE 263 AA; 28248 MW; 197C3EEB888FA242 CRC64;  
 Query Match 94.6%; Score 229; DB 4; Length 263;  
 Best Local Similarity 100.0%; Pred. No. 8e-235;  
 Matches 229; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 SGGCFWNGHLYREDQTSAPAGRLCLNWLDAQSGLASAPVSGAGNSHCNPNDEPRGPW 60  
 DB 22 SGGCFWNGHLYREDQTSAPAGRLCLNWLDAQSGLASAPVSGAGNSHCNPNDEPRGPW 81  
 QY 61 CYVSGEAGVPEKPCEDLRCPEITTSQALPAFTTEIQEASEGPGADEVQVFAPANALPARS 120  
 DB 82 CYVSGEAGVPEKPCEDLRCPEITTSQALPAFTTEIQEASEGPGADEVQVFAPANALPARS 141  
 QY 121 EAAAVQPVIGISQVRVNSKEKDLGLTGLVGLITMMVIIIAIGAGIILGYSKRGKDLK 180  
 DB 142 EAAAVQPVIGISQVRVNSKEKDLGLTGLVGLITMMVIIIAIGAGIILGYSKRGKDLK 201  
 QY 181 EQHDKVCEREMORITLPLSAFTNPTCEIVDEKTVVHTSQTVPDQEG 229  
 DB 202 EQHDKVCEREMORITLPLSAFTNPTCEIVDEKTVVHTSQTVPDQEG 250  
 RESULT 3  
 Q86YW2 PRELIMINARY; PRT; 234 AA.  
 ID Q86YW2;  
 AC Q86YW2;  
 DT 01-JUN-2003 (T-EMBLrel. 24, Created)  
 DT 01-JUN-2003 (T-EMBLrel. 24, Last sequence update)  
 DT 01-OCT-2003 (T-EMBLrel. 25, Last annotation update)  
 DE HGFL(S) protein.  
 GN HGFL.  
 OS Homo sapiens (Human)  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]

RP SEQUENCE FROM N.A.  
RA Chiang H., Chang M.;  
RL Submitted (JUL-2002) to the EMBL/GenBank/DBSJ databases.  
DR EMBL; AF528079; AAC33762.1; -.  
DR InterPro; IPR000001; Kringle.  
DR Pfam; PF00051; kringle; 1.  
DR PRINTS; PR00018; KRINGLE.  
DR SMART; SM00130; KR; 1.  
DR PROSITE; PS00021; KRINGLE\_1; 1.  
DR PROSITE; PS00070; KRINGLE\_2; 1.  
SQ SEQUENCE 234 AA; 25320 MW; C78F64DBD1B8DC0D CRC64;  
  
Query Match 72.3%; Score 175; DB 4; Length 234;  
Best Local Similarity 100.0%; Pred. No. 2.1e-177;  
Matches 175; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 1 SGCGFWNDGHLVREDQTSAPGLRCLNWLDAQSLASAPVSGAGNHSYCRNDEPRGF 60  
DB 22 SGCGFWNDGHLVREDQTSAPGLRCLNWLDAQSLASAPVSGAGNHSYCRNDEPRGF 81  
  
QY 61 CYVSGAGVPEKPCDLRCPETTSQALPAFTTEIOEASGPGADVQVFAFANALPARS 120  
DB 82 CYVSGAGVPEKPCDLRCPETTSQALPAFTTEIOEASGPGADVQVFAFANALPARS 141  
  
QY 121 EAAAVQPVIGISQVRMNSKEKDLGTGLGYLGITMMVIIAGAGIILGYSYKR 175  
DB 142 EAAAVQPVIGISQVRMNSKEKDLGTGLGYLGITMMVIIAGAGIILGYSYKR 196  
  
RESULT 4  
Q8NCJ9 PRELIMINARY; PRT; 263 AA.  
ID Q8NCJ9  
AC Q8NCJ9  
DT 01-OCT-2002 (TrEMBLrel. 22, Created)  
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)  
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)  
DE Hypothetical protein FLJ90207.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Mammary gland;  
RA Isogai T., Oca T., Nishikawa T., Hayashi K., Otsuki T., Sugiyama T.,  
RA Suzuki Y., Nagai K., Sugano S., Ishii S., Kawai-Hio Y., Saito K.,  
RA Yanamoto J., Wakamatsu A., Nakamura Y., Kojima S., Nagahari K.,  
RA Masuho Y., Ono T., Okano K., Yoshikawa Y., Aotsuka S., Sasaki N.,  
RA Hattori A., Okumura K., Iwayanagi T., Ninomiya K.;  
RA "NEDO human cDNA sequencing project";  
RL Submitted (MAR-2002) to the EMBL/GenBank/DBSJ databases.  
CC -1- SIMILARITY: CONTAINS 1 KRINGLE DOMAIN.  
DR EMBL; AK074688; BAC11140.1; -.  
DR InterPro; IPR000001; Kringle.  
DR Pfam; PF00051; kringle; 1.  
DR PRINTS; PR00018; KRINGLE.  
DR ProDom; PD000395; Kringle; 1.  
DR SMART; SM00130; KR; 1.  
DR PROSITE; PS00021; KRINGLE\_1; 1.  
DR PROSITE; PS00070; KRINGLE\_2; 1.  
KW Hypothetical protein; Glycoprotein; Kringle.  
SQ SEQUENCE 263 AA; 28104 MW; 73A9294D5142C6C3 CRC64;  
  
Query Match 41.7%; Score 101; DB 4; Length 263;  
Best Local Similarity 100.0%; Pred. No. 1.3e-99;  
Matches 101; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 142 KKDGLTGLGYLGITMMVIIAGAGIILGYSYKSKDLKEQHDQKVCERMQRITPLSA 201  
DB 163 KKDGLTGLGYLGITMMVIIAGAGIILGYSYKSKDLKEQHDQKVCERMQRITPLSA 222  
  
QY 202 FTNPTCEIVDEKTVVHTSQTFPVDPQEGSTPLMGQAGTGA 242

DB 223 FTNPTCEIVDEKTVVHTSQTFPVDPQEGSTPLMGQAGTGA 263  
  
RESULT 5  
Q81123 PRELIMINARY; PRT; 213 AA.  
ID Q81123  
AC Q81123  
DT 01-JUN-2003 (TrEMBLrel. 24, Created)  
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)  
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
DE HGFL(S) protein.  
GN HGFL.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=BAUB/c;  
RA Chiang H., Chang M.;  
RL Submitted (JUL-2002) to the EMBL/GenBank/DBSJ databases.  
DR EMBL; AF528078; AAC33761.1; -.  
DR InterPro; IPR000001; Kringle.  
DR Pfam; PF00051; kringle; 1.  
DR PRINTS; PR00018; KRINGLE.  
DR SMART; SM00130; KR; 1.  
DR PROSITE; PS00021; KRINGLE\_1; 1.  
DR PROSITE; PS00070; KRINGLE\_2; 1.  
SQ SEQUENCE 213 AA; 23011 MW; 4A5E3481025EAE97 CRC64;  
  
Query Match 10.7%; Score 26; DB 11; Length 213;  
Best Local Similarity 100.0%; Pred. No. 6.6e-19;  
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 135 VRMNSKEKKDLGTGLGYLGITMMVII 160  
DB 157 VRMNSKEKKDLGTGLGYLGITMMVII 182  
  
RESULT 6  
Q81122 PRELIMINARY; PRT; 264 AA.  
ID Q81122  
AC Q81122  
DT 01-JUN-2003 (TrEMBLrel. 24, Created)  
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)  
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
DE HGFL(L) protein.  
GN HGFL.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=BAUB/c;  
RA Chiang H., Chang M.;  
RL Submitted (JUL-2002) to the EMBL/GenBank/DBSJ databases.  
DR EMBL; AF528081; AAC33764.1; -.  
DR InterPro; IPR000001; Kringle.  
DR Pfam; PF00051; kringle; 1.  
DR PRINTS; PR00018; KRINGLE.  
DR SMART; SM00130; KR; 1.  
DR PROSITE; PS00021; KRINGLE\_1; 1.  
DR PROSITE; PS00070; KRINGLE\_2; 1.  
SQ SEQUENCE 264 AA; 28595 MW; 833EA578FEB50E34 CRC64;  
  
Query Match 10.7%; Score 26; DB 11; Length 264;  
Best Local Similarity 100.0%; Pred. No. 8e-19;  
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 135 VRMNSKEKKDLGTGLGYLGITMMVII 160  
DB 157 VRMNSKEKKDLGTGLGYLGITMMVII 182

```
RESULT 7
Q7TMJ8 PRELIMINARY; PRT; 264 AA.
ID Q7TMJ8
AC Q7TMJ8
DT 01-OCT-2003 (TrEMBLrel. 25, Created)
DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Hypothetical protein.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Breast tumor;
RX MEDLINE=22388257; PubMed=12477932;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shennan C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahy J., Helton E., Kettelman M., Madan A., Rodrigues S., Sanchez A.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
RA Krzywinski M.I., Skalska U., Smailus D.E., Schnerch A., Schein J.E.,
RA Jones S.J., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
RT and mouse cDNA sequences";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=AB; TISSUE=Body;
RA Strausberg R.;
RL Submitted (AUG-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC055920; AAH55675.1; -
KW Hypothetical protein.
SQ SEQUENCE 264 AA; 28567 MW; 833EA578FEB763A4 CRC64;

Query Match 10.7%; Score 26; DB 11; Length 264;
Best Local Similarity 100.0%; Pred. No. 8e-19;
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 135 VRNRSKKKDLGLTGLYVLTMMVII 160
Db 157 VRNRSKKKDLGLTGLYVLTMMVII 182

RESULT 8
Q7SXB3 PRELIMINARY; PRT; 263 AA.
ID Q7SXB3
AC Q7SXB3
DT 01-OCT-2003 (TrEMBLrel. 25, Created)
DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Hypothetical protein.
OS Brachydanio rerio (Zebrafish) (Danio rerio).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;
OC Cyprinidae; Danio.
OX NCBI_TaxID=7955;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=AB; TISSUE=Body;
RX MEDLINE=22388257; PubMed=12477932;
```

```
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shennan C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahy J., Helton E., Kettelman M., Madan A., Rodrigues S., Sanchez A.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
RA Krzywinski M.I., Skalska U., Smailus D.E., Schnerch A., Schein J.E.,
RA Jones S.J., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
RT and mouse cDNA sequences";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=AB; TISSUE=Body;
RA Strausberg R.;
RL Submitted (AUG-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC055675; AAH55675.1; -
KW Hypothetical protein.
SQ SEQUENCE 263 AA; 28777 MW; 8BEBCL17EC7C8A58 CRC64;

Query Match 4.5%; Score 11; DB 13; Length 263;
Best Local Similarity 100.0%; Pred. No. 0.0073;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 142 KKDGLTGLGYVL 152
Db 159 KKDGLTGLGYVL 169

RESULT 9
Q99RQ5 PRELIMINARY; PRT; 331 AA.
ID Q99RQ5
AC Q99RQ5
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Hypothetical protein SAV2372.
GN SAV2372 OR SA2162.
OS Staphylococcus aureus (strain Mu50 / ATCC 700699), and
OS Staphylococcus aureus (strain N315).
OC Bacteria; Firmicutes; Bacillales; Staphylococcus.
OX NCBI_TaxID=158878, 158879;
RN [1]
RP SEQUENCE FROM N.A.
RC SPECIES=S.aureus (strain Mu50), and S.aureus (strain N315);
RX MEDLINE=21311952; PubMed=11418146;
RA Kuroda M., Ohta T., Uchiyama I., Baba T., Yuzawa H., Kobayashi I.,
RA Cui L., Oguchi A., Aoki K.-I., Nagai Y., Iian J.-Q., Ito T.,
RA Kanamori M., Matsumaru H., Maruyama A., Murakami H., Hosoyama A.,
RA Mizutani-Ui Y., Takahashi N.K., Sawano T., Inoue R.-I., Kaito C.,
RA Sekimizu K., Hirakawa H., Kuhara S., Goto S., Yabuzaki J.,
RA Kanehisa M., Yamashita A., Oshima K., Furuya K., Yoshino C., Shiba T.,
RA Hattori M., Ogatawara N., Hayashi H., Hiratsuka K.,
RT "Whole genome sequencing of methicillin-resistant Staphylococcus
RT aureus.";
RL Lancet 357:1225-1240(2001).
CC -1- COPACTOR: FAD (BY SIMILARITY).
DR EMBL; AP003365; BAB58534.1; -
DR EMBL; AP003137; BAB43464.1; -
DR FIR; G90037; G90037.
DR DR GO; GO:0015036; F:disulfide oxidoreductase activity; IEA.
DR DR GO; GO:0006118; P:electron transport; IEA.
DR DR InterPro; IPR001327; FAD pyr redox.
DR DR InterPro; IPR00103; Pyridine_redox_2.
```

```

DR InterPro: IPR001100; Pyr_redox.
DR Pfam: PF00070; Pyr_redox; 1.
DR PRINTS: PR00368; FADPNR.
DR PRINTS: PR00411; PNDRTASEII.
DR PRINTS: PR00469; PNDRTASEII.
DR FAD; Flavoprotein; Oxidoreductase; Hypothetical protein;
KW Complete proteome.
SQ SEQUENCE 331 AA; 37034 MW; FCBB21214107FEA CRC64;

Query Match 3.7%; Score 9; DB 16; Length 331;
Best Local Similarity 100.0%; Pred. No. 1.2;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 160 IIAIGAGII 168
Db 97 IIAIGAGII 105

RESULT 10
Q8NV36 PRELIMINARY; PRT; 344 AA.
AC Q8NV36;
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE MW2294 protein.
GN MW2294.
OS Staphylococcus aureus (strain MW2).
OC Bacteria; Firmicutes; Bacillales; Staphylococcus.
OX NCBI_TaxID=196620;
RN [1]_TaxID=196620;
RP SEQUENCE FROM N.A.
RX MEDLINE=22040717; PubMed=12044378;
RA Baba T., Takeuchi F., Kuroda M., Yuzawa H., Aoki K.-I., Oguchi A.,
RA Nagai Y., Iwama N., Asano K., Naimi T., Kuroda H., Cui L.,
RA Yamamoto K., Hiramatsu K.;
RT Genome and virulence determinants of high virulence community-
RT acquired MRSA";
RL Lancet 359:1819-1827(2002).
DR ENML; AP004830; BA96159.1;
DR GO; GO:0015036; F:disulfide oxidoreductase activity; IEA.
DR GO; GO:0006118; P:electron transport; IEA.
DR InterPro: IPR000759; Adrxn_redox.
DR InterPro: IPR001327; FAD_Pyr_redox.
DR InterPro: IPR000103; Pyridine_redox_2.
DR InterPro: IPR001100; Pyr_redox.
DR Pfam: PF00070; Pyr_redox; 1.
DR PRINTS: PR00419; ADXRDTASE.
DR PRINTS: PR00368; FADPNR.
DR PRINTS: PR00411; PNDRTASEII.
DR PRINTS: PR00469; PNDRTASEII.
KW Complete proteome.
SQ SEQUENCE 344 AA; 38174 MW; 26E33154834C3C61 CRC64;

Query Match 3.7%; Score 9; DB 16; Length 344;
Best Local Similarity 100.0%; Pred. No. 1.3;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 160 IIAIGAGII 168
Db 110 IIAIGAGII 118

RESULT 11
Q800Y7 PRELIMINARY; PRT; 540 AA.
AC Q800Y7;
DT 01-JUN-2003 (TrEMBLrel. 24, Created)
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Hepatocyte growth factor activator (Fragment).
DE Hepatocyte growth factor activator (Fragment).
OS Meleagris gallopavo (Common turkey).
OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Meleagris.
OX NCBI_TaxID=9103;
RN [1]_TaxID=9103;
RP SEQUENCE FROM N.A.
RX MEDLINE=22122796; PubMed=12128863;
RA Holtsberger D.R., Becker A.E., Thurston R.J., Rice C.D.;
RA Holtsberger D.R., Becker A.E., Thurston R.J., Rice C.D.;
RT "Expression of a hepatocyte growth-factor activator protein in turkey
(Meleagris gallopavo) deferent duct epithelial cells.";
RL Comp. Biochem. Physiol. 132:769-777(2002).
RN [2]_
RP SEQUENCE FROM N.A.
RA Holtsberger D.R., Becker A.E., Thurston R.J., Rice C.D.;
RA Submitted (JAN-2003) to the EMBL/GenBank/DBJ databases.
DR ENML; AY216598; AA046038.1;
DR GO; GO:0005576; C:extracellular; IEA.
DR GO; GO:0004263; F:chymotrypsin activity; IEA.
DR GO; GO:0004295; F:trypsin activity; IEA.
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR InterPro: IPR009003; Cys_Ser_trypsin.
DR InterPro: IPR000742; EGF_2.
DR InterPro: IPR006209; EGF_like.
DR InterPro: IPR000083; Fibrinctnl.
DR InterPro: IPR000562; FN_Type_II.
DR InterPro: IPR006210; IEGF.
DR InterPro: IPR000001; Kringle.
DR InterPro: IPR001254; Peptidase_S1.
DR InterPro: IPR001314; Peptidase_S1A.
DR Pfam: PF00008; EGF; 2.
DR Pfam: PF00051; kringle; 1.
DR Pfam: PF00089; trypsin; 1.
DR PRINTS: PR00722; CHYMOTRYPSIN.
DR PRINTS: PR00018; KRINGLE.
DR ProDom; PD000395; Kringle; 1.
DR SMART; SM00181; EGF; 2.
DR SMART; SM00058; FN1; 1.
DR SMART; SM00059; FN2; 1.
DR SMART; SM00130; KR; 1.
DR SMART; SM00020; Tryp_Spc; 1.
DR PROSITE; PS00022; EGF_1; 2.
DR PROSITE; PS01186; EGF_2; 1.
DR PROSITE; PS01253; FIBRONECTIN_1; 1.
DR PROSITE; PS00021; KRINGLE_1; 1.
DR PROSITE; PS00070; KRINGLE_2; 1.
DR PROSITE; PS00240; TRYPSIN_DOM; 1.
DR PROSITE; PS00134; TRYPSIN_HIS; 1.
DR PROSITE; PS00135; TRYPSIN_SER; 1.
FT NON_TER 1
FT NON_TER 540
SQ SEQUENCE 540 AA; 61159 MW; 0BB3B4A89C0B577F CRC64;

Query Match 3.7%; Score 9; DB 13; Length 540;
Best Local Similarity 100.0%; Pred. No. 1.9;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 47 SYNCNPDED 55
Db 229 SYNCNPDED 237

RESULT 12
Q8G5B4 PRELIMINARY; PRT; 155 AA.
AC Q8G5B4;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Hypothetical protein.
GN BL1101.
OS Bifidobacterium longum.
OC Bacteria; Actinobacteria; Actinobacteridae; Bifidobacteriales;
OC Bifidobacteriaceae; Bifidobacterium.
OX NCBI_TaxID=216816;
RN [1]_

```

```

RP SEQUENCE FROM N.A.
RX STRAIN=NCC 2705;
RX MEDLINE=22294977; PubMed=12381787;
RA Schell M.A., Karmaliantzou M., Snel B., Vilanova D., Berger B.,
RA Pessi G., Zwaan M.-C., Desiere F., Borz F., Delley M.,
RA Fridmore R.D., Arigoni F.;
RA "The genome sequence of Bifidobacterium longum reflects its adaptation
RT to the human gastrointestinal tract.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:14422-14427(2002).
RA EMBL; AE014733; J024909.1; -.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 155 AA; 17838 MW; 25CC99676AEBBFF8 CRC64;

Query Match 3.3%; Score 8; DB 16; Length 155;
Best Local Similarity 100.0%; Pred. No. 6.9;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 194 RITLPLSA 201
Db 130 RITLPLSA 137
|||||

RESULT 13
Q85J73 PRELIMINARY; PRT; 345 AA.
AC Q85J73;
DT 01-JUN-2003 (TrEMBLrel. 24, Created)
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE NADH dehydrogenase subunit 2.
GN ND2.
OS Varanus rosenbergi.
OC Mitochondrion.
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Lepidosauria; Squamata; Scleroglossa; Anguilliformia; Varanidae; Varanus.
CX NCBI_TaxID=71011;
[1]
RP SEQUENCE FROM N.A.
RA Ast J.C.;
RA "Varanus biogeography, morphology, and taxonomic recommendations.";
RT Submitted (MAR-2003) to the EMBL/GenBank/DBJ databases.
RL EMBL; AY264941; AAP2711.1; -.
DR GO; GO:0005739; C:mitochondrion; IEA.
DR GO; GO:0008137; F:NADH dehydrogenase (ubiquinone) activity; IEA.
DR GO; GO:0006120; P:mitochondrial electron transport, NADH to u. . .; IEA.
DR InterPro; IPR003917; NADH dehydrogenase.
DR InterPro; IPR001750; Oxidoreductase.
DR Pfam; PF00361; oxidoreductase.
DR PRINTS; PR01436; NADH dehydrogenase.
KW Mitochondrion.
SQ SEQUENCE 345 AA; 37968 MW; B1704069029141CF CRC64;

Query Match 3.3%; Score 8; DB 8; Length 345;
Best Local Similarity 100.0%; Pred. No. 15;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 196 TLPLSAFT 203
Db 323 TLPLSAFT 330
|||||

RESULT 14
Q81WY8 PRELIMINARY; PRT; 452 AA.
AC Q81WY8;
DT 01-JUN-2003 (TrEMBLrel. 24, Created)
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Arsenical pump family protein, putative.
GN BA5462.
OS Bacillus anthracis (strain Ames).
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
CX NCBI_TaxID=198094;

[1]
RP SEQUENCE FROM N.A.
RX MEDLINE=22608414; PubMed=12721629;
RA Read T.D., Peterson S.N., Tourasse N., Baillie L.W., Paulsen I.T.,
RA Nelson K.E., Tettelin H., Fouts D.E., Eisen J.A., Gill S.R.,
RA Holtzapple E.K., Okstad O.A., Helgason E., Ristone J., Wu M.,
RA Kolonay J.F., Beanan M.J., Dodson R.J., Brinkac L.M., Gwinn M.,
RA DeBoy R.T., Macpue R., Daugherty S.C., Durkin A.S., Haft D.H.,
RA Nelson W.C., Peterson J.D., Pop M., Khouri H.M., Radune D.,
RA Barton J.L., Mahamoud Y., Jiang L., Hance I.R., Weidman J.F.,
RA Berry K.J., Plaut R.D., Wolf A.M., Watkins K.L., Nierman W.C.,
RA Hazen A., Cline R., Redmond C., Thwaite J.E., White O., Salzberg S.L.,
RA Thomsen B., Friedlander A.M., Koehler T.M., Hanna P.C., Kolsto A.-B.,
RA Fraser C.M.;
RA "The genome sequence of Bacillus anthracis Ames and comparison to
RT closely related bacteria.";
RL Nature 423:81-86(2003).
RL EMBL; AE017040; AAP29116.1; -.
DR TIGR; BA5462; -.
DR GO; GO:0016021; C:integral to membrane; IEA.
DR GO; GO:0015105; F:arsenite transporter activity; IEA.
DR InterPro; IPR000802; Ars_pump.
DR Pfam; PF02040; ArsB; 1.
DR PRINTS; PR00758; ARSENICPUMP.
KW Complete proteome.
SQ SEQUENCE 452 AA; 50862 MW; 9FB0F4FE40694E60 CRC64;

Query Match 3.3%; Score 8; DB 16; Length 452;
Best Local Similarity 100.0%; Pred. No. 19;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 162 AIGAGIIL 169
Db 33 AIGAGIIL 40
|||||

RESULT 15
Q815D6 PRELIMINARY; PRT; 452 AA.
AC Q815D6;
DT 01-JUN-2003 (TrEMBLrel. 24, Created)
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Arsenical pump membrane protein.
GN EC5226.
OS Bacillus cereus (strain ATCC 14579 / DSM 31).
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
CX NCBI_TaxID=226900;
[1]
RP SEQUENCE FROM N.A.
RX MEDLINE=22608415; PubMed=12721630;
RA Ivanova N., Sorokin A., Anderson I., Galleron N., Candelon B.,
RA Kapral V., Bhattacharya A., Reznik G., Mikhailova N., Lapidus A.,
RA Chu L., Mazur M., Goltzman E., Larsen N., D'Souza M., Walunas T.,
RA Grechkin Y., Pusch G., Haselkorn R., Fonstein M., Ehrlich S.D.,
RA Overbeek R., Kyrpides N.;
RA "Genome sequence of Bacillus cereus and comparative analysis with
RT Bacillus anthracis.";
RL Nature 423:87-91(2003).
RL EMBL; AE017014; AAP2090.1; -.
DR GO; GO:0016021; C:integral to membrane; IEA.
DR GO; GO:0015105; F:arsenite transporter activity; IEA.
DR InterPro; IPR000802; Ars_pump.
DR Pfam; PF02040; ArsB; 1.
DR PRINTS; PR00758; ARSENICPUMP.
KW Complete proteome.
SQ SEQUENCE 452 AA; 50799 MW; E921E783141368C5 CRC64;

Query Match 3.3%; Score 8; DB 16; Length 452;
Best Local Similarity 100.0%; Pred. No. 19;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 162 AIGAGIIL 169

```

```

Db      33 AIGAGIIL 40
|||||||
RESULT 16
Q82IK3 PRELIMINARY; PRT; 457 AA.
AC Q82IK3;
DT 01-JUN-2003 (TREMBlrel. 24, Created)
DT 01-JUN-2003 (TREMBlrel. 24, Last sequence update)
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
DE Putative acetyltransferase.
GN SAV3130.
OS Streptomyces avermitilis.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Streptomycineae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=33903;
RN [1]
RN SEQUENCE FROM N.A.
RC STRAIN=MA-4680 / ATCC 31267 / NCIMB 12804 / NRRL 8165;
RX MEDLINE=21477403; PubMed=11572948;
RA Omura S., Ikeda H., Ishikawa J., Hanamoto A., Takahashi C., T.,
RA Shinose M., Takahashi Y., Horikawa H., Nakazawa H., Osonoe T.,
RA Kikuchi H., Shiba T., Sakaki Y., Hattori M.;
RT "Genome sequence of an industrial microorganism Streptomyces
RT avermitilis: deducing the ability of producing secondary
RT metabolites";
RL Proc. Natl. Acad. Sci. U.S.A. 98:12215-12220(2001).
RN [2]
RN SEQUENCE FROM N.A.
RC STRAIN=MA-4680 / ATCC 31267 / NCIMB 12804 / NRRL 8165;
RX MEDLINE=22608306; PubMed=12692562;
RA Ikeda H., Ishikawa J., Hanamoto A., Shinose M., Kikuchi H., Shiba T.,
RA Sakaki Y., Hattori M., Omura S.;
RT "Complete genome sequence and comparative analysis of the industrial
RT microorganism Streptomyces avermitilis";
RL Nat. Biotechnol. 21:526-531(2003).
DR EMBL; AP005034; BAC70841.1; -
DR GO; GO:0008080; F:N-acetyltransferase activity; IEA.
DR GO; GO:0016740; F:transferase activity; IEA.
DR InterPro; IPR000182; GCS5acetyl_trans.
DR Pfam; PF00583; Acetyltransf; 1.
DR Transferrase; Complete proteome.
SQ SEQUENCE 457 AA; 46816 MW; 8A621F6CEADB4265 CRC64;

Query Match 3.3%; Score 8; DB 16; Length 457;
Best Local Similarity 100.0%; Pred. No. 19;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 63 VSAGEGVP 70
DB 352 VSAGEGVP 359
|||||||

RESULT 17
Q972D9 PRELIMINARY; PRT; 475 AA.
AC Q972D9;
DT 01-DEC-2001 (TREMBlrel. 19, Created)
DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)
DT 01-JUN-2003 (TREMBlrel. 24, Last annotation update)
DE Hypothetical protein ST1189.
GN ST1189.
OS Sulfolobus tokodaii.
OC Archaea; Crenarchaeota; Thermoprotei; Sulfolobales; Sulfolobaceae;
OC Sulfolobus.
OX NCBI_TaxID=111955;
RN [1]
RN SEQUENCE FROM N.A.
RC STRAIN=JCM 10545 / 7;
RX MEDLINE=21456156; PubMed=11572479;
RA Kawarayashi Y., Hino Y., Horikawa H., Jin-no K., Takahashi M.,
RA Sekine M., Baba S.-I., Ankaï A., Kosugi H., Hoshoya A., Fukui S.,

```

```

RA Negai Y., Nishijima K., Otsuka R., Nakazawa H., Takamiya M., Kato Y.,
RA Yoshizawa T., Tanaka T., Kudon Y., Yamazaki J., Kuchida N., Ouchi A.,
RA Aoki K.-I., Masuda S., Yanagii M., Nishimura M., Yamagishi A.,
RA Oshima T., Kikuchi H.;
RT "Complete genome sequence of an aerobic thermoadipophilic
RT Ctenarchaeon, Sulfolobus tokodaii strain 7";
RL DNA Res. 8:123-140(2001).
DR EMBL; AP000985; BAB66230.1; -
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 475 AA; 54559 MW; 9F08E45ED90374E9 CRC64;

Query Match 3.3%; Score 8; DB 17; Length 475;
Best Local Similarity 100.0%; Pred. No. 20;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 208 EIVDEKTV 215
DB 67 EIVDEKTV 74
|||||||

RESULT 18
Q37EB2 PRELIMINARY; PRT; 602 AA.
ID Q37EB2;
AC Q37EB2;
DT 01-OCT-2001 (TREMBlrel. 18, Created)
DT 01-OCT-2001 (TREMBlrel. 18, Last sequence update)
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
DE ATP-dependent Zn protease, FTSH.
GN CAC3202.
OS Clostridium acetobutylicum.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC Clostridium.
OX NCBI_TaxID=1488;
RN [1]
RN SEQUENCE FROM N.A.
RC STRAIN=ATCC 824 / DSM 792 / VGM B-1787;
RX MEDLINE=21359325; PubMed=11466286;
RA Noelling J., Breton G., Omelchenko M.V., Makarova K.S., Zeng Q.,
RA Gibson R., Lee H.M., Dubois J., Qiu D., Hitti J., Wolf Y.I.,
RA Tatusov R.L., Sabathe F., Doucette-Stamm L., Soucaille P., Daly M.J.,
RA Bennett G.N., Koonin E.V., Smith D.R.;
RT "Genome sequence and comparative analysis of the solvent-producing
RT bacterium Clostridium acetobutylicum";
RL J. Bacteriol. 183:4823-4838(2001).
DR EMBL; AE007816; AAK81138.1; -
DR FIR; G97293; G97293.
DR MEROPS; M41.009; -
DR GO; GO:0016020; C:membrane; IEA.
DR GO; GO:0005524; F:ATP binding; IEA.
DR GO; GO:0004222; F:metalloendopeptidase activity; IEA.
DR GO; GO:0000165; F:nucleotide binding; IEA.
DR GO; GO:0008233; F:peptidase activity; IEA.
DR GO; GO:0008508; F:proteolysis and peptidolysis; IEA.
DR InterPro; IPR003593; AAA_ATPase.
DR InterPro; IPR003960; AAA_sub.
DR InterPro; IPR005936; Peptidase_FtsH.
DR InterPro; IPR00642; Peptidase_M41.
DR Pfam; PF00004; AAA; 1.
DR Pfam; PF01434; Peptidase_M41; 1.
DR SMART; SM00382; AAA; 1.
DR TIGRFAMs; TIGR01241; FtsH_fam; 1.
DR PROSITE; PS00674; AAA; 1.
KW Protease; Complete proteome.
SQ SEQUENCE 602 AA; 66211 MW; 5E949997BB553A50 CRC64;

Query Match 3.3%; Score 8; DB 16; Length 602;
Best Local Similarity 100.0%; Pred. No. 25;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 63 VSAGEGVP 70
DB 213 VSAGEGVP 220
|||||||

```

```
RESULT 19
P70521 PRELIMINARY; PRT; 716 AA.
AC P70521;
DT 01-FEB-1997 (TREMBlrel. 02, Created)
DT 01-FEB-1997 (TREMBlrel. 02, Last sequence update)
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
DE Macrophage stimulating protein precursor.
GN MSP.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OC NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Liver;
RX MEDLINE=97011126; PubMed=8858136;
RA Oshiro K., Iwama A., Matsuno K., Ezaki T., Sakamoto O., Hamaguchi I.,
RA Takasu N., Suda T.
RT "Molecular cloning of Rat Macrophage-stimulating protein and its
RT involvement in the Male Reproductive System.";
RL Biochem. Biophys. Res. Commun. 227:273-280(1996).
CC -1- SIMILARITY: CONTAINS 4 KRINGLE DOMAINS.
DE EMBL; X5096; CAA64473.1; --
DR PIR; JC5061; JC5061.
DR HSP; P00747; IKNR.
DR MEROPS; S01.975; --
DR GO; GO:0005509; F:calcium ion binding; IEA.
DR GO; GO:0004263; F:chymotrypsin activity; IEA.
DR GO; GO:0008233; F:peptidase activity; IEA.
DR GO; GO:0003809; F:thrombin activity; IEA.
DR GO; GO:0004295; F:trypsin activity; IEA.
DR GO; GO:0007596; P:blood coagulation; IEA.
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR InterPro; IPR009003; Cys_Ser_trypsin.
DR InterPro; IPR000001; Kringle.
DR InterPro; IPR003014; PAN.
DR InterPro; IPR003609; Pan app.
DR InterPro; IPR001254; Peptidase_S1.
DR InterPro; IPR001314; Peptidase_S1A.
DR InterPro; IPR003966; Peptidase_S1A_pr.
DR Pfam; PF00051; Kringle; 4.
DR Pfam; PF00024; PAN; 1.
DR SMART; SM00130; KR; 4.
DR SMART; SM00473; PAN_AP; 1.
DR PROSITE; PS00020; TRYPSIN_DOM; 1.
DR PROSITE; PS00021; KRINGLE_1; 4.
DR PROSITE; PS00070; KRINGLE_2; 4.
DR PROSITE; PS00150; PROTHROMBIN.
DR ProDom; PD000395; Kringle; 4.
DR SMART; SM00130; KR; 4.
DR SMART; SM00473; PAN_AP; 1.
DR SMART; SM00020; TRYPSIN_DOM; 1.
DR PROSITE; PS00021; KRINGLE_1; 4.
DR PROSITE; PS00070; KRINGLE_2; 4.
DR PROSITE; PS00150; PROTHROMBIN.
DR GlycoProtex; GlycoProtex; Kringle; Protease; Serine protease; Signal.
FT SIGNAL; 1 31 POTENTIAL.
SQ SEQUENCE 716 AA; 80733 MW; 0687DF3EF56D921F CRC64;

Query Match 3.3%; Score 8; DB 11; Length 716;
Best Local Similarity 100.0%; Pred.No. 29;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 55 DPGPWCY 62
DB 163 DPGPWCY 170

RESULT 20
Q91XG8 PRELIMINARY; PRT; 716 AA.
AC Q91XG8;
DT 01-FEB-1997 (TREMBlrel. 02, Created)
DT 01-FEB-1997 (TREMBlrel. 02, Last sequence update)
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
DE Macrophage stimulating protein precursor.
GN MSP.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OC NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Liver;
RX MEDLINE=97011126; PubMed=8858136;
RA Oshiro K., Iwama A., Matsuno K., Ezaki T., Sakamoto O., Hamaguchi I.,
RA Takasu N., Suda T.
RT "Molecular cloning of Rat Macrophage-stimulating protein and its
RT involvement in the Male Reproductive System.";
RL Biochem. Biophys. Res. Commun. 227:273-280(1996).
CC -1- SIMILARITY: CONTAINS 4 KRINGLE DOMAINS.
DE EMBL; X5096; CAA64473.1; --
DR PIR; JC5061; JC5061.
DR HSP; P00747; IKNR.
DR MEROPS; S01.975; --
DR GO; GO:0005509; F:calcium ion binding; IEA.
DR GO; GO:0004263; F:chymotrypsin activity; IEA.
DR GO; GO:0008233; F:peptidase activity; IEA.
DR GO; GO:0003809; F:thrombin activity; IEA.
DR GO; GO:0004295; F:trypsin activity; IEA.
DR GO; GO:0007596; P:blood coagulation; IEA.
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR InterPro; IPR009003; Cys_Ser_trypsin.
DR InterPro; IPR000001; Kringle.
DR InterPro; IPR003014; PAN.
DR InterPro; IPR003609; Pan app.
DR InterPro; IPR001254; Peptidase_S1.
DR InterPro; IPR001314; Peptidase_S1A.
DR InterPro; IPR003966; Peptidase_S1A_pr.
DR Pfam; PF00051; Kringle; 4.
DR Pfam; PF00024; PAN; 1.
DR SMART; SM00130; KR; 4.
DR SMART; SM00473; PAN_AP; 1.
DR SMART; SM00020; TRYPSIN_DOM; 1.
DR PROSITE; PS00021; KRINGLE_1; 4.
DR PROSITE; PS00070; KRINGLE_2; 4.
DR PROSITE; PS00150; PROTHROMBIN.
DR GlycoProtex; GlycoProtex; Kringle; Protease; Serine protease; Signal.
FT SIGNAL; 1 31 POTENTIAL.
SQ SEQUENCE 716 AA; 80733 MW; 0687DF3EF56D921F CRC64;

Query Match 3.3%; Score 8; DB 11; Length 716;
Best Local Similarity 100.0%; Pred.No. 29;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 55 DPGPWCY 62
DB 163 DPGPWCY 170

RESULT 20
Q91XG8 PRELIMINARY; PRT; 716 AA.
AC Q91XG8;
DT 01-FEB-1997 (TREMBlrel. 02, Created)
DT 01-FEB-1997 (TREMBlrel. 02, Last sequence update)
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
DE Macrophage stimulating protein precursor.
GN MSP.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OC NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Liver;
RX MEDLINE=97011126; PubMed=8858136;
RA Oshiro K., Iwama A., Matsuno K., Ezaki T., Sakamoto O., Hamaguchi I.,
RA Takasu N., Suda T.
RT "Molecular cloning of Rat Macrophage-stimulating protein and its
RT involvement in the Male Reproductive System.";
RL Biochem. Biophys. Res. Commun. 227:273-280(1996).
CC -1- SIMILARITY: CONTAINS 4 KRINGLE DOMAINS.
DE EMBL; X5096; CAA64473.1; --
DR PIR; JC5061; JC5061.
DR HSP; P00747; IKNR.
DR MEROPS; S01.975; --
DR GO; GO:0005509; F:calcium ion binding; IEA.
DR GO; GO:0004263; F:chymotrypsin activity; IEA.
DR GO; GO:0008233; F:peptidase activity; IEA.
DR GO; GO:0003809; F:thrombin activity; IEA.
DR GO; GO:0004295; F:trypsin activity; IEA.
DR GO; GO:0007596; P:blood coagulation; IEA.
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR InterPro; IPR009003; Cys_Ser_trypsin.
DR InterPro; IPR000001; Kringle.
DR InterPro; IPR003014; PAN.
DR InterPro; IPR003609; Pan app.
DR InterPro; IPR001254; Peptidase_S1.
DR InterPro; IPR001314; Peptidase_S1A.
DR InterPro; IPR003966; Peptidase_S1A_pr.
DR Pfam; PF00051; Kringle; 4.
DR Pfam; PF00024; PAN; 1.
DR SMART; SM00130; KR; 4.
DR SMART; SM00473; PAN_AP; 1.
DR SMART; SM00020; TRYPSIN_DOM; 1.
DR PROSITE; PS00021; KRINGLE_1; 4.
DR PROSITE; PS00070; KRINGLE_2; 4.
DR PROSITE; PS00150; PROTHROMBIN.
DR GlycoProtex; GlycoProtex; Kringle; Protease; Serine protease; Signal.
FT SIGNAL; 1 31 POTENTIAL.
SQ SEQUENCE 716 AA; 80733 MW; 0687DF3EF56D921F CRC64;

Query Match 3.3%; Score 8; DB 11; Length 716;
Best Local Similarity 100.0%; Pred.No. 29;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 55 DPGPWCY 62
DB 163 DPGPWCY 170

RESULT 21
Q8PFY5 PRELIMINARY; PRT; 764 AA.
AC Q8PFY5;
DT 01-OCT-2002 (TREMBlrel. 22, Created)
DT 01-OCT-2002 (TREMBlrel. 22, Last sequence update)
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
DE Hypothetical protein XAC3837.
GN XAC3837.
OS Xanthomonas axonopodis (pv. citri).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;
OC Xanthomonadaceae; Xanthomonas.
OC NCBI_TaxID=92829;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=306 / ATCC 13902 / XV 101;
RX MEDLINE=2202145; PubMed=12024217;
RA da Silva A.C.R., Ferro J.A., Reinach F.C., Farah C.S., Furlan L.R.,
RA Quaggio R.B., Monteiro-Vitorello C.B., Van Sluys M.A., Almeida N.F.,
RA Alves L.M.C., do Amaral A.M., Bertolini M.C., Camargo L.E.A., L.P.,
RA Camarotte G., Cannavan F., Cardozo J., Chambergo F., Ciapina L.P.,
RA Ciccarelli R.M.B., Coutinho L.L., Cursino-Santos J.R., El-Dorri H.,
```

```
DT 01-DEC-2001 (TREMBlrel. 19, Created)
DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
DE Hepatocyte growth factor-like.
GN MST1 OR HGFL.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OC NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Liver;
RX MEDLINE=97011126; PubMed=8858136;
RA Oshiro K., Iwama A., Matsuno K., Ezaki T., Sakamoto O., Hamaguchi I.,
RA Takasu N., Suda T.
RT "Molecular cloning of Rat Macrophage-stimulating protein and its
RT involvement in the Male Reproductive System.";
RL Biochem. Biophys. Res. Commun. 227:273-280(1996).
CC -1- SIMILARITY: CONTAINS 4 KRINGLE DOMAINS.
DE EMBL; X5096; CAA64473.1; --
DR PIR; JC5061; JC5061.
DR HSP; P00747; IKNR.
DR MEROPS; S01.975; --
DR GO; GO:0005509; F:calcium ion binding; IEA.
DR GO; GO:0004263; F:chymotrypsin activity; IEA.
DR GO; GO:0008233; F:peptidase activity; IEA.
DR GO; GO:0003809; F:thrombin activity; IEA.
DR GO; GO:0004295; F:trypsin activity; IEA.
DR GO; GO:0007596; P:blood coagulation; IEA.
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR InterPro; IPR009003; Cys_Ser_trypsin.
DR InterPro; IPR000001; Kringle.
DR InterPro; IPR003014; PAN.
DR InterPro; IPR003609; Pan app.
DR InterPro; IPR001254; Peptidase_S1.
DR InterPro; IPR001314; Peptidase_S1A.
DR InterPro; IPR003966; Peptidase_S1A_pr.
DR Pfam; PF00051; Kringle; 4.
DR Pfam; PF00024; PAN; 1.
DR SMART; SM00130; KR; 4.
DR SMART; SM00473; PAN_AP; 1.
DR PROSITE; PS00020; TRYPSIN_DOM; 1.
DR PROSITE; PS00021; KRINGLE_1; 4.
DR PROSITE; PS00070; KRINGLE_2; 4.
DR PROSITE; PS00150; PROTHROMBIN.
DR GlycoProtex; GlycoProtex; Kringle; Protease; Serine protease.
SQ SEQUENCE 716 AA; 80693 MW; 1247C48A7D4B46D CRC64;

Query Match 3.3%; Score 8; DB 11; Length 716;
Best Local Similarity 100.0%; Pred.No. 29;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 55 DPGPWCY 62
DB 163 DPGPWCY 170

RESULT 21
Q8PFY5 PRELIMINARY; PRT; 764 AA.
AC Q8PFY5;
DT 01-OCT-2002 (TREMBlrel. 22, Created)
DT 01-OCT-2002 (TREMBlrel. 22, Last sequence update)
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
DE Hypothetical protein XAC3837.
GN XAC3837.
OS Xanthomonas axonopodis (pv. citri).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;
OC Xanthomonadaceae; Xanthomonas.
OC NCBI_TaxID=92829;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=306 / ATCC 13902 / XV 101;
RX MEDLINE=2202145; PubMed=12024217;
RA da Silva A.C.R., Ferro J.A., Reinach F.C., Farah C.S., Furlan L.R.,
RA Quaggio R.B., Monteiro-Vitorello C.B., Van Sluys M.A., Almeida N.F.,
RA Alves L.M.C., do Amaral A.M., Bertolini M.C., Camargo L.E.A., L.P.,
RA Camarotte G., Cannavan F., Cardozo J., Chambergo F., Ciapina L.P.,
RA Ciccarelli R.M.B., Coutinho L.L., Cursino-Santos J.R., El-Dorri H.,
```

```

RA Faria J.B., Ferreira A.J.S., Ferreira R.C.C., Ferro M.I.T.,
RA Formighieri E.F., Franco M.C., Greggio C.C., Gruber A.,
RA Katsuyama A.M., Kishi L.T., Leite R.P., Lemos E.G.M., Lemos M.V.F.,
RA Locali E.C., Machado M.A., Madeira A.M.B.N., Martinez-Rossi N.M.,
RA Martins E.C., Meidanis J., Menck C.F.M., Miyaki C.Y., Moon D.H.,
RA Moreira L.M., Novo M.T.M., Okura V.K., Oliveira M.C., Oliveira V.R.,
RA Pereira H.A., Rossi A., SENA J.A.D., Silva C., de Souza R.F.,
RA Spinola L.A.F., Takita M.A., Tamura R.B., Teixeira E.C., Tezza R.I.D.,
RA Trindade dos Santos M., Truffi D., Tsai S.M., White F.F.,
RA Stubal J.C., Kitajima J.P.,
RT "Comparison of the genomes of two Xanthomonas pathogens with differing
RT host specificities."
RL Nature 417:459-463 (2002).
DR EMBL: AE012033; AM38679.1; -.
DR GO: GO:0005524; F:ATP binding; IEA.
DR GO: GO:0008026; F:ATP dependent helicase activity; IEA.
DR GO: GO:0003676; F:nucleic acid binding; IEA.
DR InterPro: IPR001410; DEAD.
DR DR InterPro: IPR001650; Helicase_C.
DR Pfam: PF00271; helicase_C; 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 764 AA; 85179 MW; D3931776D1885646 CRC64;

Query Match 3.3%; Score 8; DB 16; Length 764;
Best Local Similarity 100.0%; Pred. No. 31; Mismatches 0; Indels 0; Gaps 0;
Matches 8; Conservative 0;

QY 107 VQVFAPAN 114
DB 568 VQVFAPAN 575
|||||
PRELIMINARY; PRT; 2585 AA.

RESULT 22
Q23587 Q23587 PRELIMINARY; PRT; 2585 AA.
ID AC Q23587
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Hypothetical protein ZK783.1.
GN ZK783.1.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidae;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RX MEDLINE=99069613; PubMed=9851916;
RA Waterston R.;
RT "Genome sequence of the nematode C. elegans: a platform for
RT investigating biology. The C. elegans Sequencing Consortium.";
RL Science 282:2012-2018 (1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RA Favallo A., Vaudin M.;
RT "The sequence of C. elegans cosmid ZK783.";
RL Submitted (NOV-1994) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RA Waterston R.;
RT Submitted (MAY-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL: U13646; AAC24418.2; -.
DR PIR: T34513; T34513.
DR HSP; P00736; IAPQ.
DR WormPep; ZK783.1; CE25695.
DR GO: GO:0005509; F:calcium ion binding; IEA.
DR GO: GO:000198; F:structural molecule activity; IEA.
DR GO: GO:0004295; F:trypsin activity; IEA.
DR GO: GO:0006508; P:proteolysis and peptidolysis; IEA.
DR InterPro: IPR000152; Asx_hydroxyl_S.

```

```

DR InterPro: IPR001881; EGF_Ca.
DR InterPro: IPR006209; EGF_Like.
DR InterPro: IPR001507; Endoglin/CD105.
DR InterPro: IPR006210; IEGF.
DR InterPro: IPR002049; Laminin_EGF.
DR InterPro: IPR001254; Peptidase_S1.
DR Pfam: PF00008; EGF; 15.
DR Pfam: PF00100; zona pellucida; 1.
DR PRINTS: PR00011; EGF_LAMININ.
DR SMART: SM00181; EGF; 30.
DR SMART: SM00179; EGF_CA; 16.
DR SMART: SM00241; ZP_1.
DR PROSITE: PS00010; ASX_HYDROXYL; 9.
DR PROSITE: PS00022; EGF_1; 1.
DR PROSITE: PS01186; EGF_2; 22.
DR PROSITE: PS01187; EGF_CA; 13.
DR PROSITE: PS00135; TRYPsin SER; 1.
KW Hypothetical protein; EGF_Like Domain.
SQ SEQUENCE 2585 AA; 271205 MW; 5EFD32B769CAC5B CRC64;

Query Match 3.3%; Score 8; DB 5; Length 2585;
Best Local Similarity 100.0%; Pred. No. 98;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 64 SGEAGVPE 71
DB 1269 SGEAGVPE 1276
|||||
PRELIMINARY; PRT; 37 AA.

RESULT 23
Q8WT10 Q8WT10 PRELIMINARY; PRT; 37 AA.
ID AC Q8WT10
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE ETRAMP10.1 protein (Fragment).
GN ETRAMP10.1.
OS Plasmodium falciparum (isolate 3D7).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=36329;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=3D7;
RX MEDLINE=21036610; PubMed=11163452;
RA Spielmann T., Beck H.P.;
RT "Analysis of stage-specific transcription in Plasmodium falciparum
RT reveals a set of genes exclusively transcribed in ring stage
RT parasites.";
RL Mol. Biochem. Parasitol. 111:453-458 (2000).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=3D7;
RA Spielmann T., Beck H.P.;
RT "Etramps, a new Plasmodium falciparum gene family coding for highly
RT charged membrane proteins located at the parasite-host cell
RT interface.";
RL Submitted (NOV-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: AJ420672; CAD12622.1; -.
FT NON_TER 1
FT NON_TER 37
SQ SEQUENCE 37 AA; 4035 MW; F0734FEAC69E4B58 CRC64;

Query Match 2.9%; Score 7; DB 5; Length 37;
Best Local Similarity 100.0%; Pred. No. 21;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 227 QEGSTPL 233
DB 26 QEGSTPL 32
|||||
PRELIMINARY; PRT; 37 AA.

RESULT 24

```



```

O8IK21
ID Q8IK21 PRELIMINARY; PRT; 107 AA.
AC Q8IK21
DT 01-MAR-2003 (TRENBLrel. 23, Created)
DT 01-MAR-2003 (TRENBLrel. 23, Last sequence update)
DT 01-JUN-2003 (TRENBLrel. 24, Last annotation update)
DE Early transcribed membrane protein.
GN PF10_0019.
OS Plasmodium falciparum (isolate 3D7).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporidia; Plasmodium.
OX NCBI_TaxID=36329;
RV [1]_TaxID=36329;
RP SEQUENCE FROM N.A.
RC STRAIN=3D7;
RX MEDLINE=22255705; PubMed=12369864;
RA Gardner M.J., Hall N., Fung E., White O., Berriman M., Hyman R.W.,
RA Carlton J.M., Pain A., Nelson K.E., Bowman S., Paulsen I.T., James K.,
RA Bisen J.A., Rutherford K., Salzberg S.L., Craig A., Kyes S.,
RA Chan M.-S., Nene V., Shallow S.J., Sub B., Peterson J., Angiuoli S.,
RA Perlea M., Allen J., Selengut J., Haft D., Mather M.W., Vaidya A.B.,
RA Martin D.M.A., Fairlamb A.H., Fraunholz M.J., Roos D.S., Ralph S.A.,
RA McFadden G.I., Cummings L.M., Subramanian G.M., Mungall C.,
RA Venter J.C., Carucci D.J., Hoffman S.L., Newbold C., Davis R.W.,
RA Fraser C.M., Barrell B.;
RT "Genome sequence of the human malaria parasite Plasmodium
RT falciparum";
RL Nature 419:498-511 (2002).
DR EMBL; AB014829; AAN35217.1; -.
DR InterPro; IPR006389; ETAMP.
DR TIGRPFAM; TIGR01495; ETAMP; 1.
SQ SEQUENCE 107 AA; 11267 MW; A808F8DE6C01923A CRC64;

Query Match 2.9%; Score 7; DB 5; Length 107;
Best Local Similarity 100.0%; Pred. No. 56;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 227 QEGSTPL 233
Db 96 QEGSTPL 102

RESULT 25
O8TK63
ID Q8TK63 PRELIMINARY; PRT; 112 AA.
AC Q8TK63
DT 01-JUN-2002 (TRENBLrel. 21, Created)
DT 01-JUN-2002 (TRENBLrel. 21, Last sequence update)
DT 01-JUN-2003 (TRENBLrel. 24, Last annotation update)
DE Cobalt ABC transporter, solute-binding protein.
GN CBIN OR MA3553.
OS Methanosarcina acetivorans.
OC Archaea; Euryarchaeota; Euryarchaeota orders incertae sedis;
OC Methanosarcinales; Methanosarcinaceae; Methanosarcina.
CX NCBI_TaxID=2214;
RV [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C2A / ATCC 35395 / DSM 2834;
RX MEDLINE=21929760; PubMed=11932238;
RA Galagan J.E., Nusbaum C., Roy A., Endrizzi M.G., MacDonald P.,
RA FitzHugh W., Calvo S., Engels R., Smirnov S., Atnoor D., Brown A.,
RA Allen N., Naylor J., Stange-Thomann N., Dearellano K., Johnson R.,
RA Linton L., McSwan P., McKernan K., Talamas J., Tirrell A., Ye W.,
RA Zimmer A., Barber R.D., Cann I., Graham D.E., Grahame D.A., Guss A.M.,
RA Hedderich R., Ingram-Smith C., Kuettnner H.C., Krzycki J.A.,
RA Leigh J.A., Li W., Liu J., Mukhopadhyay B., Reeve J.N., Smith K.,
RA Springer T.A., Unayam L.A., White O., White R.H., de Macario E.C.,
RA Ferry J.G., Jarrell K.F., Jing H., Macario A.J.L., Paulsen I.,
RA Pritchett M., Sowers K.R., Swanson R.V., Zinder S.H., Lander E.,
RA Metcalf W.W., Birren B.;
RT "The genome of Methanosarcina acetivorans reveals extensive metabolic
RT and physiological diversity.";
RL Genome Res. 12:532-542 (2002).
DR EMBL; AB011064; AAM06915.1; -.

DR GO; GO:0016020; C:membrane; IEA.
DR GO; GO:0015087; P:cobalt ion transporter activity; IEA.
DR GO; GO:0006824; P:cobalt ion transport; IEA.
DR GO; GO:0009236; P:vitamin B12 biosynthesis; IEA.
DR InterPro; IPR003705; CblN.
DR Pfam; PF02553; CblN; 1.
KW Complete proteome.
SQ SEQUENCE 112 AA; 12244 MW; C98FA33AA9468E55 CRC64;

Query Match 2.9%; Score 7; DB 17; Length 112;
Best Local Similarity 100.0%; Pred. No. 59;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 162 AIGAGII 168
Db 75 AIGAGII 81

RESULT 26
Q8PZM8
ID Q8PZM8 PRELIMINARY; PRT; 113 AA.
AC Q8PZM8
DT 01-OCT-2002 (TRENBLrel. 22, Created)
DT 01-OCT-2002 (TRENBLrel. 22, Last sequence update)
DT 01-JUN-2003 (TRENBLrel. 24, Last annotation update)
DE Cobalt transport protein.
GN CBIN OR MM0484.
OS Methanosarcina mazei (Methanosarcina frisia).
OC Archaea; Euryarchaeota; Euryarchaeota orders incertae sedis;
OC Methanosarcinales; Methanosarcinaceae; Methanosarcina.
OX NCBI_TaxID=2209;
RV [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Goel / GoI / ATCC BAA-199 / DSM 3647 / OCM 88;
RX MEDLINE=22120827; PubMed=12125824;
RA Deppenmeier U., Johann A., Hartsch T., Merkl R., Schmitz R.A.,
RA Martinez-Arias R., Henne A., Wieser A., Baumer S., Jacobl C.,
RA Brueggemann H., Lienard T., Christmann A., Boemcke M., Steckel S.,
RA Bhattacharya A., Lykidis A., Overbeek R., Klenk H.-P., Gunsalus R.P.,
RA Fritz H.-J., Gottschalk G.;
RT "The genome of Methanosarcina mazei: evidence for lateral gene
RT transfer between Bacteria and Archaea.";
RL J. Mol. Microbiol. Biotechnol. 4:453-461 (2002).
DR EMBL; AB013272; AAM30160.1; -.
DR GO; GO:0016020; C:membrane; IEA.
DR GO; GO:0015087; P:cobalt ion transporter activity; IEA.
DR GO; GO:0006824; P:cobalt ion transport; IEA.
DR GO; GO:0009236; P:vitamin B12 biosynthesis; IEA.
DR InterPro; IPR003705; CblN.
DR Pfam; PF02553; CblN; 1.
KW Complete proteome.
SQ SEQUENCE 113 AA; 12428 MW; 16AE084D3D311480 CRC64;

Query Match 2.9%; Score 7; DB 17; Length 113;
Best Local Similarity 100.0%; Pred. No. 59;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 162 AIGAGII 168
Db 75 AIGAGII 81

RESULT 27
O13544
ID O13544 PRELIMINARY; PRT; 120 AA.
AC O13544
DT 01-JAN-1998 (TRENBLrel. 05, Created)
DT 01-JAN-1998 (TRENBLrel. 05, Last sequence update)
DT 01-JUN-2003 (TRENBLrel. 24, Last annotation update)
DE YLR302CP.
GN YLR302CP.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;

```

OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.  
 ON NCBI\_TaxID=4932;  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=97313267; PubMed=9169871;  
 RA Johnston M., Hillier L., Riles L., Albertmann K., Andre B., Ansoerge W.,  
 RA Benes V., Bruckner M., Dillus H., Dubois E., Dusterhoft A.,  
 RA Entian K.D., Floeth M., Goffeau A., Hebling U., Heumann K.,  
 RA Heuss-Neitzel D., Hilbert H., Hilger F., Kleine K., Kottler P.,  
 RA Louis E.J., Messenguy F., Mewes H.W., Miosga T., Mostl D.,  
 RA Muller-Auer S., Nentwich U., Obermaier B., Piravandi B., Pohl T.M.,  
 RA Portetelle D., Purnelle B., Reckmann S., Rieger M., Rinke M., Rose M.,  
 RA Scharfe M., Scherens B., Scholler P., Schwager C., Schwarz S.,  
 RA Underwood A.P., Urrestarazu L.A., Vandenbol M., Verhasselt P.,  
 RA Vierendeels F., Voet M., Volckert G., Voss H., Wambutt R., Wedler E.,  
 RA Wedler H., Zimmermann F.K., Zollner A., Hanl J., Hohlseil J.D.; XII.;  
 RL "The nucleotide sequence of Saccharomyces cerevisiae chromosome XII.";  
 RN Nature 387:0-0(0).  
 RP SEQUENCE FROM N.A.  
 RA Pauley A.; (DEC-1994) to the EMBL/GenBank/DBJ databases.  
 RL Submitted (NOV-1994) to the EMBL/GenBank/DBJ databases.  
 RP SEQUENCE FROM N.A.  
 RA Waterston R.;  
 RL Submitted (NOV-1994) to the EMBL/GenBank/DBJ databases.  
 RP SEQUENCE FROM N.A.  
 RA Cherry J.M.;  
 RL Submitted (AUG-1997) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; U17243; AAB67353.1; -;  
 DR PIR; S69308; S69308.  
 DR SGD; S0004293; YLR302C.  
 SQ SEQUENCE 120 AA; 14090 MW; 879002561C7DAC67 CRC64;  
 Query Match 2.9%; Score 7; DB 3; Length 120;  
 Best Local Similarity 100.0%; Pred. No. 63;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 208 EIVDEKT 214  
 Db 25 EIVDEKT 31  
 RESULT 28  
 O42428 PRELIMINARY; PRT; 129 AA.  
 AC O42428;  
 DT 01-JAN-1998 (TrEMBLrel. 05, Created)  
 DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)  
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
 DE Tyrosine hydroxylase (Fragment).  
 OS TH.  
 ON Lates calcarifer (Barramundi).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;  
 OC Acanthomorpha; Acanthopterygii; Percormorpha; Perciformes; Percoidae;  
 OC Centropomidae; Lates.  
 OC NCBI\_TaxID=8187;  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=barramundi;  
 RA Collet C., Candy J., Sara V.;  
 RT "Tyrosine hydroxylase and insulin-like growth factor-II but not  
 RT insulin are adjacent in the teleost species barramundi (Lates  
 RT calcarifer).";  
 RL Submitted (JUN-1997) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AF007942; AAB64194.1; -;  
 DR HSSP; P04177; 1TOH.  
 DR GO; GO:0005506; F:iron ion binding; IEA.  
 DR GO; GO:0004497; F:monooxygenase activity; IEA.  
 DR GO; GO:0009072; P:aromatic amino acid family metabolism; IEA.  
 DR InterPro; IPR001273; Aaa\_hydroxylase.

DR Pfam; PF00351; biotpterin\_H; 1.  
 DR PRINTS; PR00372; FYWHYDRYLASE.  
 FT NON\_TER 1  
 SQ SEQUENCE 129 AA; 14678 MW; 1AE29C7530AB4D78 CRC64;  
 Query Match 2.9%; Score 7; DB 13; Length 129;  
 Best Local Similarity 100.0%; Pred. No. 67;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 121 EAAAVOP 127  
 Db 47 EAAAVOP 53  
 RESULT 29  
 Q94DN5 PRELIMINARY; PRT; 135 AA.  
 AC Q94DN5;  
 DT 01-DEC-2001 (TrEMBLrel. 19, Created)  
 DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)  
 DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)  
 DE P0483G10.8 protein (P0401G10.30 protein).  
 GN P0483G10.8 OR P0401G10.30.  
 OS Oryza sativa (Rice), and  
 OS Oryza sativa (japonica cultivar-group).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
 OC Ehrhartoideae; Oryzaceae; Oryza.  
 OC NCBI\_TaxID=4530; 39947;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=cv. Nipponbare;  
 RA Sasaki T., Matsumoto T., Yamamoto K.;  
 RT "Oryza sativa nipponbare (GA3) genomic DNA, chromosome 1, PAC  
 RT clone:P0483G10.";  
 RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=cv. Nipponbare;  
 RA Sasaki T., Matsumoto T., Yamamoto K.;  
 RT "Oryza sativa (japonica cultivar-group) genomic DNA, chromosome 1, PAC  
 RT clone:P0401G10.";  
 RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AP003263; BAB3614.1; -;  
 DR EMBL; AP003238; BAB89022.1; -;  
 DR Gramene; Q94DN5; -;  
 SQ SEQUENCE 135 AA; 13916 MW; 4B06F9BA65ABAD95 CRC64;  
 Query Match 2.9%; Score 7; DB 10; Length 135;  
 Best Local Similarity 100.0%; Pred. No. 70;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 118 ARSEAAA 124  
 Db 23 ARSEAAA 29  
 RESULT 30  
 Q8TLC3 PRELIMINARY; PRT; 139 AA.  
 AC Q8TLC3;  
 DT 01-JUN-2002 (TrEMBLrel. 21, Created)  
 DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)  
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
 DE Predicted protein.  
 GN MA3113.  
 OS Methanosarcina acetivorans.  
 OC Archaea; Euryarchaeota; Euryarchaeota orders incertae sedis;  
 OC Methanosarcinales; Methanosarcinaceae; Methanosarcina.  
 OC NCBI\_TaxID=2214;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=C2A / ATCC 35395 / DSM 2834;

```

RX MEDLINE=21929760; PubMed=11932236;
RA Galagan J.E., Nusbaum C., Roy A., Endrizzi M.G., Macdonald P.,
RA Fitzhugh W., Calvo S., Engels R., Smirnov S., Atnoor D., Brown A.,
RA Allen N., Naylor J., Stange-Thomann N., Dearellano K., Johnson R.,
RA Lincon L., McEwan P., McKernan K., Talamas J., Tittell A., Ye W.,
RA Zimmer A., Barber R.D., Cann I., Graham D.E., Grubbs D.A., Guss A.M.,
RA Hedderich R., Ingram-Smith C., Kuettnner H.C., Krzycki J.A.,
RA Leigh J.A., Li W., Liu J., Mukhopadhyay B., Reeve J.N., Smith K.,
RA Ferry J.G., Jarrell K.F., Jing H., Macario A.J.L., de Macario E.C.,
RA Pritchett M., Sowers K.R., Swanson R.V., Zinder S.H., Lander E.,
RA Metcalf W.W., Birren B.;
RT "The genome of Methanococcus acetivorans reveals extensive metabolic
RT and physiological diversity."
RL Genome Res. 12:532-542(2002).
DR EMBL; AEO11014; AAM06496.1; -.
DR InterPro; IPR009002; FMN_binding.
SQ SEQUENCE 139 AA; 15457 MW; 37D003E9BA0C8754 CRC64;

Query Match 2.9%; Score 7; DB 17; Length 139;
Best Local Similarity 100.0%; Pred. No. 72;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 211 DEXTVVV 217
DB 39 DEXTVVV 45

RESULT 31
Q55181 PRELIMINARY; PRT; 144 AA.
AC Q55181;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Hypothetical protein slr0491.
OS Synechocystis sp. (strain PCC 6803).
OC Bacteria; Cyanobacteria; Chroococcales; Synechocystis.
OX NCBI_TaxID=1148;
RN Synechocystis sp. (strain PCC 6803).
RA Tabata S.;
RC STRAIN=PCC 6803;
RL Submitted (AUG-1995) to the EMBL/GenBank/DBJ databases.
[2]
SEQUENCE FROM N.A.
MEDLINE=96127529; PubMed=8590279;
RA Kaneko T., Tanaka A., Sato S., Kotani H., Sazuka T., Miyajima N.,
RA Sugita M., Tabata S.;
RT "Sequence analysis of the genome of the unicellular cyanobacterium
RT Synechocystis sp. strain PCC6803. I. Sequence features in the 1 Mb
RT region from map positions 64% to 92% of the genome."
RL DNA Res. 2:153-166(1995).
[3]
SEQUENCE FROM N.A.
MEDLINE=97061201; PubMed=905231;
RA Kaneko T., Sato S., Kotani H., Tanaka A., Asamizu E., Nakamura Y.,
RA Miyajima N., Hiroseawa M., Sugita M., Sasamoto S., Kimura T.,
RA Hoshino T., Matsuno A., Muraki A., Nakazaki N., Naruo K., Okumura S.,
RA Shimo S., Takeuchi C., Wada T., Watanabe A., Yamada M., Yasuda M.,
RA Tabata S.;
RT "Sequence analysis of the genome of the unicellular cyanobacterium
RT Synechocystis sp. strain PCC6803. II. Sequence determination of the
RT entire genome and assignment of potential protein-coding regions."
RL DNA Res. 3:109-136(1996).
DR EMBL; D64001; BAA10321.1; -.
DR PIR; S74403; S74403.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 144 AA; 16726 MW; B6993A5507435742 CRC64;

Query Match 2.9%; Score 7; DB 16; Length 144;
Best Local Similarity 100.0%; Pred. No. 76;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 154 ITMMVII 160
DB 122 ITMMVII 128

RESULT 33
Q94XU1 PRELIMINARY; PRT; 147 AA.
AC Q94XU1;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE NADH dehydrogenase subunit 5 (fragment).
GN ND5.
OS Crotales durissus (tropical rattlesnake).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Lepidosauria; Squamata; Scleroglossa; Serpentes; Colubroidea;
OC Viperidae; Crotalinae; Crotalus.
OX NCBI_TaxID=8731;
RN [1]
SEQUENCE FROM N.A.
RA Murphy R.W., Fu J., Lathrop A., Feltham J.V., Kovac V.;
RT "Phylogeny of the rattlesnakes (Crotalus and Sistrurus) inferred from
RT sequences of five mitochondrial DNA genes."
RL Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF259211; AAL25699.1; -.
DR GO; GO:0005739; C:mitochondrion; IEA.
DR GO; GO:0008137; F:NADH dehydrogenase (ubiquinone) activity; IEA.
DR GO; GO:0006120; P:mitochondrial electron transport, NADH to u. . .; IEA.
DR InterPro; IPR003916; NADHdb_oxred5.
DR InterPro; IPR001516; Oxidored_g1_N; 1.
DR Pfam; PF00662; oxidored_g1_N; 1.
DR PRINTS; PR01434; NADHGHNASES.
KW Mitochondrion.
FT NON TER 147
SQ SEQUENCE 147 AA; 16924 MW; 925734D12CD0BB45 CRC64;

Query Match 2.9%; Score 7; DB 8; Length 147;
Best Local Similarity 100.0%; Pred. No. 76;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 154 ITMMVII 160
DB 122 ITMMVII 128

RESULT 33
Q94XU1 PRELIMINARY; PRT; 147 AA.
AC Q94XU1;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE NADH dehydrogenase subunit 5 (fragment).
GN ND5.
OS Crotales durissus (tropical rattlesnake).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Lepidosauria; Squamata; Scleroglossa; Serpentes; Colubroidea;
OC Viperidae; Crotalinae; Crotalus.
OX NCBI_TaxID=8731;
RN [1]
SEQUENCE FROM N.A.
RA Murphy R.W., Fu J., Lathrop A., Feltham J.V., Kovac V.;
RT "Phylogeny of the rattlesnakes (Crotalus and Sistrurus) inferred from
RT sequences of five mitochondrial DNA genes."
RL Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF259212; AAL25700.1; -.
DR GO; GO:0005739; C:mitochondrion; IEA.

```

DR GO: GO:0008137; F:NADH dehydrogenase (ubiquinone) activity; IEA.  
DR GO: GO:0006120; P:mitochondrial electron transport, NADH to u. . . ; IEA.  
DR InterPro: IPR003916; NADHox oxdeds.  
DR Pfam: PF00662; Oxidored g1\_N; 1.  
DR PRINTS: PR01434; NADHGHNA5E5.  
KW Mitochondrion.  
FT NON\_TER 147 147  
SQ SEQUENCE 147 AA; 16924 MW; 925734D12CD0BB45 CRC64;  
  
Query Match 2.9%; Score 7; DB 8; Length 147;  
Best Local Similarity 100.0%; Pred. No. 76;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 154 ITM0VII 160  
DB 122 ITM0VII 128  
  
RESULT 34  
Q94XU3 PRELIMINARY; PRT; 147 AA.  
AC Q94XU3; 147 AA.  
DT 01-DEC-2001 (TrEMBLrel. 19, Created)  
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)  
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
DE NADH dehydrogenase subunit 5 (fragment).  
GN ND5.  
OS Crotaulus unicolor (Aruba island rattlesnake).  
OG Mitochondrion.  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Lepidosauria; Squamata; Scleroglossa; Serpentes; Colubroidea;  
OC Viperidae; Crotalinae; Crotalus.  
OX NCBI\_TaxID=125874;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Murphy R.W., Fu J., Lathrop A., Feltham J.V., Kovac V.;  
RT "Phylogeny of the rattlesnakes (Crotalus and Sistrurus) inferred from  
sequences of five mitochondrial DNA genes."  
RL Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.  
DR EMBL: AF25210; AAL25698.1; -.  
DR GO: GO:0005733; C:mitochondrion; IEA.  
DR GO: GO:0008137; F:NADH dehydrogenase (ubiquinone) activity; IEA.  
DR GO: GO:0006120; P:mitochondrial electron transport, NADH to u. . . ; IEA.  
DR InterPro: IPR003916; NADHox oxdeds.  
DR Pfam: PF00662; Oxidored g1\_N; 1.  
DR PRINTS: PR01434; NADHGHNA5E5.  
KW Mitochondrion.  
FT NON\_TER 147 147  
SQ SEQUENCE 147 AA; 16781 MW; 7166899AB226EB84 CRC64;  
  
Query Match 2.9%; Score 7; DB 8; Length 147;  
Best Local Similarity 100.0%; Pred. No. 76;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 154 ITM0VII 160  
DB 122 ITM0VII 128  
  
RESULT 35  
Q82K32 PRELIMINARY; PRT; 150 AA.  
AC Q82K32; 150 AA.  
DT 01-MAR-2002 (TrEMBLrel. 20, Created)  
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)  
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)  
DE Putative cytoplasmic protein.  
GN YJGK OR STM468.  
OS Salmonella typhimurium.  
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
OC Enterobacteriaceae; Salmonella.

OX NCBI\_TaxID=602;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=LT2 / SGSC1412 / ATCC 700720;  
RX MEDLINE=21534948; PubMed=11677609;  
RA McClelland M., Sanderson K.E., Spieth J., Clifton S.W., Latreille P.,  
Courtney L., Porwollik S., Ali J., Dante M., Du F., Hou S., Layman D.,  
Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Mulvaney E.,  
Ryan E., Sun H., Florea L., Miller W., Stoneking T., Nhan M.,  
Waterston R., Wilson R.K.;  
RT "Complete genome sequence of Salmonella enterica serovar Typhimurium  
LT2";  
RL Nature 413:852-856(2001).  
DR EMBL: AE008909; AAL23287.1; -.  
DR InterPro: IPR004375; Cons\_hypoth22.  
DR Pfam: PF04074; DUF386; 1.  
DR TIGRFAMs: TIGR00022; TIGR00022; 1.  
KW Hypothetical protein; Complete proteome.  
SQ SEQUENCE 150 AA; 16710 MW; 4405D0D6935EF3D CRC64;  
  
Query Match 2.9%; Score 7; DB 16; Length 150;  
Best Local Similarity 100.0%; Pred. No. 78;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 210 VDEXTVV 216  
DB 107 VDEXTVV 113  
  
RESULT 36  
Q82124 PRELIMINARY; PRT; 150 AA.  
ID Q82124  
AC Q82124;  
DT 01-MAR-2002 (TrEMBLrel. 20, Created)  
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)  
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)  
DE Hypothetical protein SRY4806.  
GN STY4806 OR T4502.  
OS Salmonella typhi.  
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
OC Enterobacteriaceae; Salmonella.  
OX NCBI\_TaxID=601;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=CT18;  
RX MEDLINE=21534947; PubMed=11677608;  
RA Parkhill J., Dougan G., James K.D., Thomson N.R., Pickard D., Wain J.,  
Churcher C., Mungall K.L., Bentley S.D., Holden M.T.G., Sebahia M.,  
Baker S., Basham D., Brooks K., Chillingworth T., Connor P.,  
Cronin A., Davis P., Davies R.M., Dowd L., White N., Farrar J.,  
Feltwell T., Hamlin N., Haque A., Hien T.T., Holroyd S., Jagels K.,  
Krogh A., Larsen T.S., Leather S., Moule S., O'Gaora P., Parry C.,  
Quail M., Rutherford K., Simmonds M., Skelton J., Stevens K.,  
Whitehead S., Barrall B.G.;  
RT "Complete genome sequence of a multiple drug resistant Salmonella  
enterica serovar Typhi CT18";  
RL Nature 413:848-852(2001).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=TY2 / ATCC 700931;  
RX MEDLINE=22531367; PubMed=12644504;  
RA Deng W., Liou S.-R., Plunkett G. III, Mayhew G.F., Rose D.J.,  
Burland V., Kodoyianni V., Schwartz D.C., Blattner F.R.;  
RT "Comparative genomics of Salmonella enterica serovar Typhi strains Ty2  
and CT18";  
RL J. Bacteriol. 185:2330-2337(2003).  
DR EMBL: AL627283; CAD06928.1; -.  
DR EMBL: AB016849; AAC071949.1; -.  
DR InterPro: IPR004375; Cons\_hypoth22.  
DR Pfam: PF04074; DUF386; 1.  
DR TIGRFAMs: TIGR00022; TIGR00022; 1.  
KW Hypothetical protein; Complete proteome.  
SQ SEQUENCE 150 AA; 16638 MW; 4D7383B5375E1DED CRC64;

```

Query Match      2.9%; Score 7; DB 16; Length 150;
Best Local Similarity 100.0%; Pred. No. 78;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      210 VDEKTV 216
      |||||
DB      107 VDEKTV 113

RESULT 37
Q8KUE6      PRELIMINARY; PRT; 151 AA.
AC      Q8KUE6;
DT      01-OCT-2002 (TREMBlrel. 22, Created)
DT      01-OCT-2002 (TREMBlrel. 22, Last sequence update)
DT      01-JUN-2003 (TREMBlrel. 24, Last annotation update)
DE      Lrp-like regulator.
GN      LRP.
OS      Corynebacterium glutamicum (Brevibacterium flavum).
OC      Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC      Corynebacterineae; Corynebacteriaceae; Corynebacterium.
OX      NCBI_TaxID=1718;
RN      [1]
RP      SEQUENCE FROM N.A.
RX      MEDLINE=22077265; PubMed=12081967;
RA      Kennerknecht N., Sahm H., Yen M.R., Patek M., Saier M.H. Jr.,
RA      Eggeling L.;
RT      "Export of L-Isoleucine from Corynebacterium glutamicum: a Two-Gene-
RT      Encoded Member of a New Translocator Family.";
RL      J. Bacteriol. 184:3947-3956(2002).
CC      -!- SIMILARITY: BELONGS TO THE ASNC FAMILY OF TRANSCRIPTIONAL
CC      REGULATORS.
DR      EMBL; AF545053; AAM46687.1; -.
DR      GO; GO:0005622; C:intracellular; IEA.
DR      GO; GO:0003700; P:transcription factor activity; IEA.
DR      GO; GO:0006350; P:transcription; IEA.
DR      GO; GO:0006350; P:transcription; IEA.
DR      InterPro; IPR000485; HTH_AsnC_lrp.
DR      Pfam; PF01037; ASNC_trans_reg; 1.
DR      PRINTS; PR00033; HTHASNC.
DR      SMART; SMO0344; HTH_ASN_C; 1.
DR      DNA-binding; Transcription; Transcription regulation.
KW      DNA-binding; Transcription; Transcription regulation.
SQ      SEQUENCE 151 AA; 16702 MW; 76B21368D31CB8A3 CRC64;

Query Match      2.9%; Score 7; DB 2; Length 151;
Best Local Similarity 100.0%; Pred. No. 78;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      166 GILGYS 172
      |||||
DB      47 GILGYS 53

RESULT 38
Q8NTP9      PRELIMINARY; PRT; 154 AA.
AC      Q8NTP9;
DT      01-OCT-2002 (TREMBlrel. 22, Created)
DT      01-OCT-2002 (TREMBlrel. 22, Last sequence update)
DT      01-JUN-2003 (TREMBlrel. 24, Last annotation update)
DE      Transcriptional regulators.
GN      CGL0257.
OS      Corynebacterium glutamicum (Brevibacterium flavum).
OC      Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC      Corynebacterineae; Corynebacteriaceae; Corynebacterium.
OX      NCBI_TaxID=1718;
RN      [1]
RP      SEQUENCE FROM N.A.
RX      STRAIN=ATCC 13032 / DSM 20300 / NCIB 10025;
RA      Nakagawa S.;
RT      "Complete genomic sequence of Corynebacterium glutamicum ATCC 13032.";
RL      Submitted (MAY-2002) to the EMBL/GenBank/DBJ databases.

```

---

```

DR      EMBL; AF005274; BAB97650.1; -.
DR      GO; GO:0005622; C:intracellular; IEA.
DR      GO; GO:0003700; P:transcription factor activity; IEA.
DR      GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.
DR      InterPro; IPR000485; HTH_AsnC_lrp.
DR      Pfam; PF01037; ASNC_trans_reg; 1.
DR      PRINTS; PR00033; HTHASNC.
KW      Complete proteome.
SQ      SEQUENCE 154 AA; 17124 MW; 06B603D9B6DD790E CRC64;

Query Match      2.9%; Score 7; DB 16; Length 154;
Best Local Similarity 100.0%; Pred. No. 80;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      166 GILGYS 172
      |||||
DB      50 GILGYS 56

RESULT 39
Q9C892      PRELIMINARY; PRT; 164 AA.
AC      Q9C892;
DT      01-JUN-2001 (TREMBlrel. 17, Created)
DT      01-JUN-2001 (TREMBlrel. 17, Last sequence update)
DT      01-OCT-2003 (TREMBlrel. 25, Last annotation update)
DE      Hypothetical protein.
GN      F7A10.6.
OS      Arabidopsis thaliana (Mouse-ear cress).
OC      Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC      Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC      eurosids II; Brassicales; Brassicaceae; Arabidopsids.
OX      NCBI_TaxID=3702;
RN      [1]
RP      SEQUENCE FROM N.A.
RX      STRAIN=ecv. Columbia.
RX      MEDLINE=21016719; PubMed=11130712;
RA      Theologis A., Ecker J.R., Palm C.J., Federspiel N.A., Kaul S.,
RA      White O., Alonso J., Altati H., Araujo R., Bowman C.L., Brooks S.Y.,
RA      Buehler E., Chan A., Chao Q., Chen H., Cheuk R.F., Chin C.W.,
RA      Chung M.K., Conn L., Conway A.B., Conway A.R., Creasy T.H., Dewar K.,
RA      Dunn P., Egu P., Feldblyum T.V., Feng J.-D., Fong B., Fujii C.Y.,
RA      Gill J.B., Goldsmith A.D., Haas B., Hansen N.F., Hughes B., Huizar L.,
RA      Hunter J.L., Jenkins J., Johnson-Hopson C., Khan S., Khaykin E.,
RA      Kim C.J., Koo H.L., Kremenetskaia I., Kurtz D.B., Kvan A., Lam B.,
RA      Langin-Hooper S., Lee A., Lee J.M., Lenz C.A., Li J.H., Li Y.-P.,
RA      Lin X., Liu S.X., Liu Z.A., Luros J.S., Maiti R., Marziani A.,
RA      Militscher J., Miranda M., Nguyen M., Nierman W.C., Osborne B.I.,
RA      Pai G., Peterson J., Pham P.K., Rizzio M., Rooney T., Rowley D.,
RA      Sakano H., Salzberg S.L., Schwartz J.R., Shinn P., Southwick A.M.,
RA      Sun H., Tallon L.J., Tambunga G., Toriumi M.J., Town C.D.,
RA      Utterback T., Van Aken S., Vaysberg M., Vysotskaia V.S., Walker M.,
RA      Wu D., Yu G., Fraser C.M., Venter J.C., Davis R.W.;
RT      "Sequence and analysis of chromosome 1 of the plant Arabidopsis
RT      thaliana.";
RL      Nature 408:816-820(2000).
DR      EMBL; AC027034; AAG51579.1; -.
DR      PIR; A96594; A96594.
KW      Hypothetical protein.
SQ      SEQUENCE 164 AA; 18498 MW; 97C054B6B82B80BA CRC64;

Query Match      2.9%; Score 7; DB 10; Length 164;
Best Local Similarity 100.0%; Pred. No. 84;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      142 KDLGLT 148
      |||||
DB      145 KDLGLT 151

RESULT 40
Q89SV2      PRELIMINARY; PRT; 176 AA.
ID      Q89SV2

```

Q895V2;  
 DT 01-JUN-2003 (TRENBLrel. 24, Created)  
 DT 01-JUN-2003 (TRENBLrel. 24, Last sequence update)  
 DT 01-JUN-2003 (TRENBLrel. 24, Last annotation update)  
 DE Membrane associated protein.  
 GN CTC01165  
 OS Clostridium tetani.  
 OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;  
 OC Clostridium.  
 OX NCBI\_TaxID=1513;  
 [1]\_TaxID=1513;  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Massachusetts / E88;  
 RX MEDLINE=22457253; PubMed=12552129;  
 RA Bruggemann H., Baumer S., Fricke W.F., Wierzer A., Liesegang H.,  
 RA Decker I., Herzberg C., Martinez-Arias R., Merkl R., Henne A.,  
 RA Gottschalk G.;  
 RT "The genome sequence of Clostridium tetani, the causative agent of  
 RT tetanus disease."  
 RL Proc. Natl. Acad. Sci. U.S.A. 100:1316-1321(2003).  
 DR EMBL; A5015940; AAO35738.1; --  
 KW Complete proteome.  
 SQ SEQUENCE 176 AA; 20553 MW; DCS327DDFE6794EF CRC64;  
 Query Match 2.9%; Score 7; DB 16; Length 176;  
 Best Local Similarity 100.0%; Pred. No. 90;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 162 AIGAGII 168  
 |||||  
 DB 42 AIGAGII 48  
 |||||  
 RESULT 41  
 Q89QF0  
 ID Q89QF0 PRELIMINARY; PRT; 185 AA.  
 AC Q89QF0;  
 DT 01-JUN-2003 (TRENBLrel. 24, Created)  
 DT 01-JUN-2003 (TRENBLrel. 24, Last sequence update)  
 DT 01-OCT-2003 (TRENBLrel. 25, Last annotation update)  
 DE Transcriptional regulatory protein.  
 GN BLJ3378  
 OS Bradyrhizobium japonicum.  
 OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;  
 OC Bradyrhizobiaceae; Bradyrhizobium.  
 OX NCBI\_TaxID=375;  
 [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=USDA 110.  
 RX MEDLINE=22484998; PubMed=12597275;  
 RA Kaneko T., Nakamura Y., Sato S., Minamisawa K., Uchiumi T.,  
 RA Sasamoto S., Watanabe A., Idesawa K., Iriguchi M., Kawashima K.,  
 RA Kohara M., Matsumoto M., Shingo S., Tsuruoka H., Wada T., Yamada M.,  
 RA Tabata S.;  
 RT "Complete genomic sequence of nitrogen-fixing symbiotic bacterium  
 RT Bradyrhizobium japonicum USDA110."  
 RL DNA Res. 9:189-197(2002).  
 DR EMBL; AP005946; BAC48443.1; --  
 DR GO; GO:0005622; C:intracellular; IEA.  
 DR GO; GO:0003700; P:transcription factor activity; IEA.  
 DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.  
 DR InterPro; IPR000485; HTH\_AsnC\_lrp.  
 DR Pfam; PF01037; ASNC\_trans\_reg7.1.  
 DR PRINTS; PR00033; HTFASNC.  
 KW Complete proteome.  
 SQ SEQUENCE 185 AA; 20233 MW; 75C762FD19159846 CRC64;  
 Query Match 2.9%; Score 7; DB 16; Length 185;  
 Best Local Similarity 100.0%; Pred. No. 94;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 37 SAPVSGA 43  
 |||||

Db 10 SAPVSGA 16  
 RESULT 42  
 Q44013  
 ID Q44013 PRELIMINARY; PRT; 192 AA.  
 AC Q44013;  
 DT 01-NOV-1996 (TRENBLrel. 01, Created)  
 DT 01-JAN-1999 (TRENBLrel. 09, Last sequence update)  
 DT 01-OCT-2003 (TRENBLrel. 25, Last annotation update)  
 DE Granule-associated protein (PHASIN).  
 GN PHAP.  
 OS Alcaligenes eutrophus (Ralstonia eutropha).  
 OC Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;  
 OC Burkholderiaceae; Ralstonia.  
 OX NCBI\_TaxID=510;  
 [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=H16;  
 RX "DNA and protein sequencing indicate that the polyhydroxyalkanoic  
 RT acid granules in Alcaligenes eutrophus."  
 RT Submitted (AUG-1998) to the EMBL/GenBank/DBJ databases.  
 RL [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=H16;  
 RX Hanley S.Z., Pappin D.J.C., White A.J., Elborough K.M., Slabas A.R.;  
 RT "DNA and protein sequencing indicate that the carboxy-terminal region  
 RT of phasins implicated in polyalkanoic acid granule binding does not  
 RT form part of the in vivo protein."  
 RL FEBS Lett. 0:0-0(1999).  
 RL [3]  
 RP SEQUENCE FROM N.A.  
 RC York G.M., Stubbe J., Sinskey A.J.;  
 RX "The Ralstonia eutropha phasin Phap promotes synthesis of  
 RT polyhydroxybutyrate throughout the period of polyhydroxybutyrate  
 RT production and across a range of cultivation conditions."  
 RL Submitted (OCT-2000) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; X85729; CAA59734.1; --  
 DR EMBL; AF079155; AAC78327.1; --  
 DR EMBL; AF314206; AAG33635.1; --  
 DR InterPro; IPR001778; POA\_allergen\_C.  
 DR PRINTS; PR00833; POAALLERGEN.  
 SQ SEQUENCE 192 AA; 19966 MW; 3A69C094ED9A3ECE CRC64;  
 Query Match 2.9%; Score 7; DB 2; Length 192;  
 Best Local Similarity 100.0%; Pred. No. 98;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 122 AAAVQPV 128  
 |||||  
 DB 69 AAAVQPV 75  
 |||||  
 RESULT 43  
 Q98M02  
 ID Q98M02 PRELIMINARY; PRT; 197 AA.  
 AC Q98M02;  
 DT 01-OCT-2001 (TRENBLrel. 18, Created)  
 DT 01-OCT-2001 (TRENBLrel. 18, Last sequence update)  
 DT 01-OCT-2003 (TRENBLrel. 25, Last annotation update)  
 DE Hypothetical protein ml10757.  
 GN MLL0797.  
 OS Rhizobium loti (Mesorhizobium loti).  
 OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;  
 OC Phyllobacteriaceae; Mesorhizobium.  
 OX NCBI\_TaxID=381;  
 [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=MAFF303099;  
 RX MEDLINE=21082930; PubMed=11214968;  
 RA Kaneko T., Nakamura Y., Sato S., Asamizu E., Kato T., Sasamoto S.,  
 RA Watanabe A., Idesawa K., Ishikawa A., Kawashima K., Kimura T.,  
 RA Kishida Y., Kiyokawa C., Kohara M., Matsumoto M., Matsuno A.,

RA Mochizuki Y., Nakayama S., Nakazaki N., Shimpo S., Sugimoto M.,  
 RA Takeuchi C., Yamada M., Tabata S.;  
 RA "Complete genome structure of the nitrogen-fixing symbiotic bacterium  
 RT Mesorhizobium loti";  
 RL DNA Res. 7:331-338 (2000).  
 DR EMBL; AF002995; BAB48311.1; -;  
 DR InterPro; IPR005986; DUF330.  
 DR InterPro; IPR000437; Prok\_lipoprot S.  
 DR Pfam; PF03886; DUF330; 1.  
 DR PROSITE; PS00013; PROKAR\_LIPOPROTEIN; 1.  
 KW Hypothetical protein, Complete proteome.  
 SQ SEQUENCE 197 AA; 20709 MW; 95829BD677C6BB74 CRC64;  
 Query Match 2.9%; Score 7; DB 16; Length 197;  
 Best Local Similarity 100.0%; Pred. No. 1.e+02;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 36 ASAPVSG 42  
 DB 162 ASAPVSG 168  
 RESULT 44  
 ID Q93IG1 PRELIMINARY; PRT; 209 AA.  
 AC Q93IG1  
 DT 01-DEC-2001 (TREMBLrel. 19, Created)  
 DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)  
 DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)  
 DE Hypothetical protein.  
 OS Propionibacterium freudenreichii shermanii.  
 OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
 OC Propionibacteriinae; Propionibacteriaceae; Propionibacterium.  
 OX NCBI\_TaxID=1752;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Reusser C.A., Huang K., Scott A.I.;  
 RT "Cobalamin biosynthesis in Propionibacterium freudenreichii  
 RT (shermanii): Isolation and characterization of 16 vitamin B12 genes."  
 RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AY033236; AAK67498.1; -;  
 DR EMBL; AY033236; AAK67498.1; -;  
 KW Hypothetical protein.  
 SQ SEQUENCE 209 AA; 21906 MW; DL653DB38370DD98 CRC64;  
 Query Match 2.9%; Score 7; DB 2; Length 209;  
 Best Local Similarity 100.0%; Pred. No. 1.e+02;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 83 TTSQALP 89  
 DB 123 TTSQALP 129  
 RESULT 45  
 ID Q9RT90 PRELIMINARY; PRT; 222 AA.  
 AC Q9RT90  
 DT 01-MAY-2000 (TREMBLrel. 13, Created)  
 DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)  
 DT 01-JUN-2003 (TREMBLrel. 24, Last annotation update)  
 DE Hypothetical protein DR1875.  
 GN DR1875  
 OS Deinococcus radiodurans.  
 OC Bacteria; Deinococcus-Thermus; Deinococci; Deinococcales;  
 OC Deinococcaceae; Deinococcus.  
 OX NCBI\_TaxID=1239;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=R1 / ATCC 13939 / DSM 20539 / NCIB 9279;  
 RX MEDLINE=20036896; PubMed=10567266;  
 RA White O., Eisen J.A., Heidelberg J.F., Hickey E.K., Peterson J.D.,  
 RA Dodson R.J., Haft D.H., Gwinn M.L., Nelson W.C., Richardson D.L.,  
 RA McFait K.S., Qin H., Jiang L., Pamphile W., Crosby M., Shen M.,

RA Vamathevan J.J., Lam P., McDonald L., Utterback T., Zalewski C.,  
 RA Makarova K.S., Aravind L., Daly M.J., Minton K.W., Fleischmann R.D.,  
 RA Ketchum K.A., Nelson K.E., Salzberg S., Smith H.O., Venter J.C.,  
 RA Fraser C.M.;  
 RT "Genome sequence of the radioresistant bacterium Deinococcus  
 RT radiodurans R1";  
 RL Science 286:1571-1577 (1999).  
 DR EMBL; AS002027; AAF11429.1; -;  
 DR FIC; C75342; C75342.  
 DR TIGR; DR1875; -;  
 DR InterPro; IPR003740; DUF161.  
 DR Pfam; PF02588; DUF161; 2.  
 KW Hypothetical protein, Complete proteome.  
 SQ SEQUENCE 222 AA; 22884 MW; 8E542E41B016EC09 CRC64;  
 Query Match 2.9%; Score 7; DB 16; Length 222;  
 Best Local Similarity 100.0%; Pred. No. 1.e+02;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 35 LASAPVS 41  
 DB 2 LASAPVS 8  
 RESULT 46  
 ID P96683 PRELIMINARY; PRT; 226 AA.  
 AC P96683  
 DT 01-MAY-1997 (TREMBLrel. 03, Created)  
 DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)  
 DT 01-JUN-2003 (TREMBLrel. 24, Last annotation update)  
 DE YDFP protein.  
 GN YDFP.  
 OS Bacillus subtilis.  
 OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.  
 OX NCBI\_TaxID=1423;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=168;  
 RX MEDLINE=98000887; PubMed=9341680;  
 RA Beloin C., Ayora S., Exley R., Hirschbein L., Ogasawara N.,  
 RA Kasahara Y., Alonso J.C., Le Hegarat F.,  
 RT "Characterization of an lrp-like (lrpC) gene from Bacillus subtilis."  
 RL Mol. Gen. Genet. 256:63-71 (1997).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=168;  
 RX MEDLINE=98044033; PubMed=9384377;  
 RA Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,  
 RA Azevedo V., Bertero M.G., Bessieres P., Bolotin A., Borchert S.,  
 RA Boriss R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S.,  
 RA Brouillet S., Bruschi C.V., Caldwell B., Capuano V., Carter N.M.,  
 RA Choi S.K., Codani J.J., Connerton I.F., Cummings N.J., Daniel R.A.,  
 RA Denizot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emerson P.T.,  
 RA Entian K.D., Brington J., Fabret C., Ferrari E., Foulger D.,  
 RA Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,  
 RA Ghim S.Y., Glaser P., Goffeau A., Golightly E.J., Grandi G.,  
 RA Guisepi G., Guy B.J., Haga K., Haiech J., Harwood C.R., Henaut A.,  
 RA Hilbert H., Holsappel S., Hosono S., Hullo M.F., Itaya M., Jones L.,  
 RA Joris B., Karamata D., Kashara Y., Klaerr-Blanchard M., Klein C.,  
 RA Kobayashi Y., Koetter P., Koningsstein G., Krogh S., Kumano M.,  
 RA Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V.,  
 RA Lee S.M., Levine A., Liu H., Masuda S., Mauei C., Medigue C.,  
 RA Medina N., Mellado R.P., Mizuno M., Moesti D., Nakai S., Noback M.,  
 RA Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H.,  
 RA Parro V., Pohl T.M., Portetelle D., Porwollik S., Prescott A.M.,  
 RA Prescan E., Pujić P., Purnelle B., Rapoport G., Rey M., Reynolds S.,  
 RA Rieger M., Rivolta C., Rocha E., Roche B., Rose M., Sadate Y.,  
 RA Sato T., Scanlan E., Schleich S., Schroeter R., Scroffone F.,  
 RA Sekiguchi J., Sekowska A., Seror S.J., Serrin P., Shin B.S., Soldo B.,  
 RA Sorokin A., Taconi E., Takagi T., Takahashi H., Takemaru K.,  
 RA Takeuchi M., Tamakoshi A., Tanaka T., Terpstra P., Tognoni A.,  
 RA Tosato V., Uchiyama S., Vandenbol M., Vannier F., Vassarotti A.,

RA Viari A., Wambutt R., Wedler E., Wedler H., Weitzenecker T., Yata K.,  
RA Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,  
RA Yoshida K., Yoshikawa H.F., Zumbstein E., Yoshikawa H., Danchin A.,  
RT "The complete genome sequence of the Gram-positive bacterium Bacillus  
RT subtilis";  
RL Nature 390:249-256(1997).  
RN [3]  
RP SEQUENCE FROM N.A.  
RC STRAIN=168;  
RA Kunst F., Ogasawara N., Yoshikawa H., Danchin A.,  
RL Submitted (NOV-1997) to the EMBL/GenBank/DBJ databases.  
CC -!- SIMILARITY: BELONGS TO THE ARSR FAMILY OF TRANSCRIPTIONAL  
CC REGULATORS.  
DR EMBL; AB001488; BAA19373.1; -;  
DR EMBL; 299106; CAB12346.1; -;  
DR EIR; D69780; D69780.  
DR HSSP; P30340; ISMT.  
DR GO; GO:0005622; C:intracellular; IEA.  
DR GO; GO:0003700; F:transcription factor activity; IEA.  
DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.  
DR InterPro; IPR001845; HTH\_ArsR.  
DR Pfam; PF01022; HTH 5; 1  
DR PRINTS; PR00778; HTHARS.  
DR SMART; SM00416; HTH\_ArsR; 1.  
KW DNA-binding; Transcription regulation; Complete proteome.  
SQ SEQUENCE 226 AA; 25617 MW; 96C44D638D2E1ABD CRC64;

Query Match 2.9%; Score 7; DB 16; Length 226;  
Best Local Similarity 100.0%; Pred. No. 1.1e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 139 SKEKDL 145  
Db 99 SKEKDL 105  
|||||

RESULT 47

Q9UUB2 PRELIMINARY; PRT; 229 AA.  
ID Q9UUB2  
AC Q9UUB2;  
DT 01-MAY-2000 (TrEMBLrel. 13, Created)  
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)  
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
DE Hypothetical protein.  
DE SPEC009.12C.  
GN SPEC009.12C.  
OS Schizosaccharomyces pombe (Fission yeast).  
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;  
OC Schizosaccharomycetales; Schizosaccharomycetaceae;  
OC Schizosaccharomycetes.  
OX NCBI\_TaxID=4896;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=972h-;  
RA Lyne M.H., Rajandream M.A., Barrell B.G., Chillingworth T.,  
RA Churcher C.M.,  
RL Submitted (AUG-1999) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AL109822; CAB52614.1; -;  
DR EIR; T40439; T40439.  
DR GeneDB; SPombe; SPB0409.12C; -;  
DR KW Hypothetical protein.  
SQ SEQUENCE 229 AA; 26714 MW; 937C8D170A5A063 CRC64;

Query Match 2.9%; Score 7; DB 3; Length 229;  
Best Local Similarity 100.0%; Pred. No. 1.2e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 195 ITPLSA 201  
Db 192 ITPLSA 198  
|||||

RESULT 48

Q88IT4

ID Q88IT4 PRELIMINARY; PRT; 233 AA.  
AC Q88IT4;  
DT 01-JUN-2003 (TrEMBLrel. 24, Created)  
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)  
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)  
DE Membrane protein, putative.  
GN PP2915.  
OS Pseudomonas putida (strain KT2440).  
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;  
OC Pseudomonadaceae; Pseudomonas.  
OX NCBI\_TaxID=160488;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=22423080; PubMed=12534463;  
RA Nelson K.E., Weinel C., Paulsen I.T., Dodson R.J., Hilbert H.,  
RA Martins dos Santos V.A.P., Fouts D.E., Gill S.R., Fop M., Holmes M.,  
RA Brinkac L., Beanan M., DeBoy R.T., Daugherty S., Kolonay J.,  
RA Madupu R., Nelson W., White O., Peterson J., Khouri H., Hance I.,  
RA Chris Lee P., Holtzapple E., Scanlan D., Tran K., Moazzes A.,  
RA Uterback T., Rizzo M., Lee K., Kosack D., Moestl D., Wedler H.,  
RA Lauber J., Stjepandic D., Hehseisel J., Straetz M., Helm S.,  
RA Kiewitz C., Eisen J., Timmis K.N., Duesterhoeft A., Tuemmli B.,  
RA Fraser C.M.;  
RT "Complete genome sequence and comparative analysis of the  
RT metabolically versatile Pseudomonas putida KT2440.";  
RL Environ. Microbiol. 4:799-808(2002).  
DR EMBL; AE016784; AAN68523.1; -;  
DR TIGR; PP2915; -;  
KW Complete proteome.  
SQ SEQUENCE 233 AA; 24531 MW; 4001F1CFC528DBF3 CRC64;

Query Match 2.9%; Score 7; DB 16; Length 233;  
Best Local Similarity 100.0%; Pred. No. 1.2e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 114 NALPARS 120  
Db 4 NALPARS 10  
|||||

RESULT 49

Q891Q5 PRELIMINARY; PRT; 254 AA.  
ID Q891Q5  
AC Q891Q5;  
DT 01-JUN-2003 (TrEMBLrel. 24, Created)  
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)  
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)  
DE Putative structural protein/integral membrane protein.  
DE C1C02313.  
GN C1C02313.  
OS Clostridium tetani.  
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;  
OC Clostridium.  
OX NCBI\_TaxID=1513;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Massachusetts / E88;  
RX MEDLINE=22457253; PubMed=2552129;  
RA Brueggemann H., Baeumer S., Fricke W.F., Wierer A., Liesegang H.,  
RA Decker I., Herzberg C., Martinez-Arias R., Merkl R., Henne A.,  
RA Gottschalk G.;  
RT "The genome sequence of Clostridium tetani, the causative agent of  
RT tetanus disease.";  
RL Proc. Natl. Acad. Sci. U.S.A. 100:1316-1321(2003).  
DR EMBL; AE015944; AAO36790.1; -;  
KW Complete proteome.  
SQ SEQUENCE 254 AA; 27912 MW; E8D8B9C4E8EB70A3 CRC64;

Query Match 2.9%; Score 7; DB 16; Length 254;  
Best Local Similarity 100.0%; Pred. No. 1.3e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 161 IAGAGI 167  
Db 161 IAGAGI 167  
|||||



Query Match 2.9%; Score 7; DB 16; Length 272;  
 Best Local Similarity 100.0%; Pred. No. 1.4e+02; Indels 0; Gaps 0;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 192 MORITLP 198  
 DB 90 MQEITLP 96

RESULT 52  
 Q9ASMO PRELIMINARY; PRT; 273 AA.  
 ID Q9ASMO  
 AC Q9ASMO  
 DT 01-JUN-2001 (TREMELrel. 17, Created)  
 DT 01-JUN-2001 (TREMELrel. 17, Last sequence update)  
 DT 01-JUN-2003 (TREMELrel. 24, Last annotation update)  
 DE Phage SPO1 DNA polymerase-related protein.  
 GN CCI1333.  
 OS Caulobacter crescentus.  
 OC Bacteria; Proteobacteria; Alphaproteobacteria; Caulobacteriales;  
 OC Caulobacteraceae; Caulobacter.  
 NCBI\_TaxID=155892;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=ATCC 19089 / CB15;  
 RX MEDLINE=21173698; PubMed=11259647;  
 RA Nierman W.C., Feldblyum T.V., Laub M.T., Paulsen I.T., Nelson K.E.,  
 RA Eisen J., Heidelberg J.F., Alley M.R.K., Ohta N., Maddock J.R.,  
 RA Pockock I., Nelson W.C., Newton A., Stephens C., Phadke N.D., Ely B.,  
 RA DeBoy R.T., Dodson R.J., Durkin A.S., Gwinn M.L., Haft D.H.,  
 RA Kolonay J.F., Smit J., Craven M.B., Khouri H., Shetty J., Berry K.,  
 RA Utterback T., Tran K., Wolf A., Vamathevan J., Ermolaeva M., White O.,  
 RA Salzberg S.L., Venter J.C., Shapiro L., Fraser C.M.;  
 RT "Complete genome sequence of Caulobacter crescentus.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 98:4136-4141(2001).  
 DR EMBL; AE005808; AAC23314.1; -;  
 DR FIC; F87414; F87414.  
 DR TIGR; CC1333; -;  
 DR InterPro; IPR005273; SPO1polNrel.  
 DR InterPro; IPR005122; UDNA\_glycosylaseSF.  
 DR Pfam; PF03167; UDG; 1.  
 DR TIGRFAMs; TIGR00758; SPO1polNrel; 1.  
 KW Complete proteome.  
 SQ SEQUENCE 273 AA; 29214 MW; CE54307DB1568194 CRC64;

Query Match 2.9%; Score 7; DB 16; Length 273;  
 Best Local Similarity 100.0%; Pred. No. 1.4e+02; Indels 0; Gaps 0;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 100 EGPGADE 106  
 DB 123 EGPGADE 129

RESULT 53  
 Q9VQU4 PRELIMINARY; PRT; 282 AA.  
 ID Q9VQU4  
 AC Q9VQU4  
 DT 01-MAY-2000 (TREMELrel. 13, Created)  
 DT 01-MAY-2000 (TREMELrel. 13, Last sequence update)  
 DT 01-OCT-2003 (TREMELrel. 25, Last annotation update)  
 DE CG3410 protein.  
 GN LECTIN-24A OR CG3410.  
 OS Drosophila melanogaster (Fruit fly).  
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
 OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 OC Ephydroidea; Drosophilidae; Drosophila.  
 OC NCBI\_TaxID=7227;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Berkeley;  
 RX MEDLINE=20196006; PubMed=10731132;  
 ADAMS M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,

Db 237 TAIGAGI 243

RESULT 50  
 Q8BH49 PRELIMINARY; PRT; 266 AA.  
 ID Q8BH49  
 AC Q8BH49  
 DT 01-WAR-2003 (TREMELrel. 23, Created)  
 DT 01-WAR-2003 (TREMELrel. 23, Last sequence update)  
 DT 01-JUN-2003 (TREMELrel. 24, Last annotation update)  
 DE Hypothetical pleckstrin homology.  
 GN A230106M1SRK.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=C57BL/6J; TISSUE=Cecum, and Hypothalamus;  
 RX MEDLINE=22354683; PubMed=12466851;  
 RA The PANTOM Consortium,  
 RA the RIKEN Genome Exploration Research Group Phase I & II Team;  
 RT "Analysis of the mouse transcriptome based on functional annotation of  
 60,770 full-length cDNAs.";  
 RL Nature 420:563-573(2002).  
 DR EMBL; AK033618; BAC28394.1; -;  
 DR EMBL; AK039192; BAC30272.1; -;  
 DR MGD; MGI:2442708; A230106M1SRK.  
 DR InterPro; IPR001849; PH.  
 DR Pfam; PF00169; PH; 1.  
 DR SMART; SM00233; PH; 1.  
 DR PROSITE; PS00003; PH\_DOMAIN; 1.  
 DR Hypothetical protein.  
 KW Hypothetical protein.  
 SQ SEQUENCE 266 AA; 29218 MW; 4D0C3B71A5D2394F CRC64;

Query Match 2.9%; Score 7; DB 11; Length 266;  
 Best Local Similarity 100.0%; Pred. No. 1.3e+02;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 112 PANALPA 118  
 DB 135 PANALPA 141

RESULT 51  
 Q8XNE0 PRELIMINARY; PRT; 272 AA.  
 ID Q8XNE0  
 AC Q8XNE0  
 DT 01-WAR-2002 (TREMELrel. 20, Created)  
 DT 01-WAR-2002 (TREMELrel. 20, Last sequence update)  
 DT 01-JUN-2003 (TREMELrel. 24, Last annotation update)  
 DE Purine nucleoside phosphorylase.  
 GN DEOD OR CPE0398.  
 OS Clostridium perfringens.  
 OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;  
 OC Clostridium.  
 OC NCBI\_TaxID=1502;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=13 / Type A;  
 RX MEDLINE=21664373; PubMed=11792842; Oshima K., Yamashita A.,  
 RA Shiba T., Ogasawara N., Hattori M., Kuhara S., Hayashi H.;  
 RT "Complete genome sequence of Clostridium perfringens, an anaerobic  
 flesh-eater.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 99:996-1001(2002).  
 DR EMBL; AP003186; BAB80104.1; -;  
 DR GO; GO:0004645; F.phosphorylase activity; IEA.  
 DR InterPro; IPR001369; Mtap\_PNP.  
 DR Pfam; PF00896; Mtap\_PNP; 1.  
 KW Complete proteome.  
 SQ SEQUENCE 272 AA; 30064 MW; F54F42F21AF6695F CRC64;

RT "Genome sequence of the nematode C. elegans: a platform for  
investigating biology. The C. elegans Sequencing Consortium."  
Science 282:2012-2018 (1998).  
[2]  
RN SEQUENCE FROM N.A.  
RC STRAIN=Bristol N2;  
RA Waterston R.;  
RL Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.  
DR EMBL: AC024811; AAF60775.1; -  
DR WormPep; Y54E10BL.2; CE28434.  
DR InterPro; IPR008160; Collagen.  
DR InterPro; IPR002486; Col cuticle\_N.  
DR Pfam; PF01391; Collagen; 2.  
DR Pfam; PF01484; Col cuticle\_N; 1.  
SQ SEQUENCE 285 AA; 27871 MW; 5568D837E02B3815 CRC64;  
Query Match 2.9%; Score 7; DB 5; Length 285;  
Best Local Similarity 100.0%; Pred. No. 1.4e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 235 GQAGTPG 241  
Db 256 GQAGTPG 262  
RESULT 55  
Q99WR0 PRELIMINARY; PRT; 286 AA.  
AC Q99WR0;  
DT 01-JUN-2001 (TrEMBLrel. 17, Created)  
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)  
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
DE Hypothetical protein SAV0316.  
GN SAV0316 OR SA0305.  
OS Staphylococcus aureus (strain Mu50 / ATCC 700699), and  
OS Staphylococcus aureus (strain N315).  
OC Bacteria; Firmicutes; Bacillales; Staphylococcus.  
OX NCBI\_TaxID=158878, 158879;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC SPECIES=S.aureus (strain Mu50), and S.aureus (strain N315);  
RX MEDLINE=21311952; PubMed=11418146;  
RA Kuroda M., Ohta T., Uchiyama I., Baba T., Yuzawa H., Kobayashi I.,  
RA Cui L., Oguchi A., Aoki K.-I., Nagai Y., Lian J.-Q., Ito T.,  
RA Kanamori M., Matsumaru H., Maruyama A., Murakami H., Hosoyama A.,  
RA Mizutani-Ui Y., Takahashi N.K., Sawano T., Inoue R.-I., Kaito C.,  
RA Sekimizu K., Hirakawa H., Kuhara S., Goto S., Yabuzaki J.,  
RA Kanehisa M., Yamashita A., Oshima K., Furuya K., Yoshino C., Shiba T.,  
RA Rattori M., Ogasawara N., Hayashi H., Hiramatsu K.;  
RT "Whole genome sequencing of methicillin-resistant Staphylococcus  
aureus."  
RL Lancet 357:1225-1240(2001).  
DR EMBL: AP003359; BAB56478.1; -  
DR EMBL: AP003130; BAB1529.1; -  
DR PIR; F89796; F89796.  
DR InterPro; IPR006000; ROK.  
DR Pfam; PF00480; ROK; 1.  
DR Hypothetical protein; Complete proteome.  
KW SEQUENCE 286 AA; 31685 MW; 6EB4CBAA77291209 CRC64;  
Query Match 2.9%; Score 7; DB 16; Length 286;  
Best Local Similarity 100.0%; Pred. No. 1.4e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 126 QPVIGIS 132  
Db 59 QPVIGIS 65  
RESULT 56  
Q8NYC6 PRELIMINARY; PRT; 286 AA.  
ID Q8NYC6  
AC Q8NYC6;

RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.P.,  
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,  
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,  
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,  
RA Abril J.P., Abayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,  
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,  
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotter P.,  
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,  
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,  
RA de Pablo B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
RA Durbin K.J., Evangelista C.C., Ferraz C., Farriera S., Fleischmann W.,  
RA Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,  
RA Glodok A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,  
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,  
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,  
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,  
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleeb J.M.,  
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
RA Swirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissbach J.,  
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu H.O.,  
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
RT "The genome sequence of Drosophila melanogaster."  
RL Science 287:2185-2195(2000).  
DR EMBL: AE003579; AAF51070.1; -  
DR FlyBase; FBgn0040104; lectin-24A.  
DR GO; GO:0005529; F:sugar binding; IEA.  
DR InterPro; IPR001304; Lectin\_C.  
DR Pfam; PF00059; lectin\_C; 1.  
DR SMART; SW00034; CLECT; 1.  
DR PROSITE; PS00041; C-TYPE-LECTIN\_2; 1.  
SQ SEQUENCE 282 AA; 32320 MW; COD833D11E544791 CRC64;  
Query Match 2.9%; Score 7; DB 5; Length 282;  
Best Local Similarity 100.0%; Pred. No. 1.4e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 26 LNWLDQAQ 32  
Db 178 LNWLDQAQ 184  
RESULT 54  
Q9N3D7 PRELIMINARY; PRT; 285 AA.  
AC Q9N3D7;  
DT 01-OCT-2000 (TrEMBLrel. 15, Created)  
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)  
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)  
DE Y54E10BL.2 protein.  
GN Y54E10BL.2  
OS Caenorhabditis elegans.  
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;  
OC Rhabditidae; Pelodierinae; Caenorhabditis.  
OX NCBI\_TaxID=6239;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Bristol N2;  
RX MEDLINE=99069613; PubMed=9851916;  
None;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0

QY 33 SGLASAP 39  
Db 89 SGLASAP 95

RESULT 58  
Q8PJQ9 PRELIMINARY; PRT; 288 AA.

ID Q8PJQ9  
AC Q8PJQ9;  
DT 01-OCT-2002 (T-EMBLrel. 22, Created)  
DT 01-OCT-2002 (T-EMBLrel. 22, Last sequence update)  
DT 01-JUN-2003 (T-EMBLrel. 24, Last annotation update)  
DE Polyamine transport protein.  
GN POTH OR XAC2471.  
OS Xanthomonas axonopodis (pv. citri).  
OC Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;  
OC Xanthomonadaceae; Xanthomonas.  
CX NCBI\_TaxID=92829;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=306 / ATCC 13902 / XV 101;  
RX MEDLINE=2020214; PubMed=12024217;  
RA da Silva A.C.R., Ferri J.A., Reinach F.C., Farah C.S., Furian L.R.,  
RA Quaggio R.B., Monteiro-Vitorello C.B., Van Sluys M.A., Almeida N.F.,  
RA Alves L.M.C., do Amaral A.M., Bertolini M.C., Camargo L.E.A.,  
RA Cimarotte G., Cannavan F., Cardoso J., Chambergo F., Ciapina L.P.,  
RA Cicarelli R.M.B., Coutinho L.L., Cursino-Santos J.R., El-Dorry H.,  
RA Faria J.B., Ferreira A.J.S., Ferreira R.C.C., Ferro M.I.T.,  
RA Formighieri E.F., Franco M.C., Greggio C.C., Gruber A.,  
RA Katsuyama A.F., Kisai L.T., Leite R.P., Lemos E.G.M., Lemos M.V.F.,  
RA Locali E.C., Machado M.A., Madeira A.M.B.N., Martinez-Rossi N.M.,  
RA Martins E.C., Meidanis J., Menck C.F.M., Miyaki C.Y., Moon D.H.,  
RA Moreira L.M., Novo M.T.M., Okura V.K., Oliveira M.C., Oliveira V.R.,  
RA Pereira H.A., Rossi A., Sena J.A.D., Silva C., de Souza R.F.,  
RA Spinola L.A.F., Takita M.A., Tamura R.E., Teixeira E.C., Tezza R.I.D.,  
RA Trindade dos Santos M., Truffi D., Tsai S.M., White F.F.,  
RA Secubal J.C., Kitajima J.P.;  
RT "Comparison of the genomes of two Xanthomonas pathogens with differing  
RT host specificities";  
RL Nature 417:459-463(2002).  
DR EMBL; AE011884; AAM37322.1; -  
DR GO; GO:0016020; C:membrane; IEA.  
DR GO; GO:0005215; F:transporter activity; IEA.  
DR GO; GO:0006810; P:transport; IEA.  
DR InterPro; IPR000515; BPD\_transp.  
DR Pfam; PF00528; BPD\_transp; I.  
DR PROSITE; PS00402; BPD\_TRANSF\_INN\_MEMBER; 1.  
KW Complete proteome.  
SQ SEQUENCE 288 AA; 31804 MW; 7716CC72128B94B5 CRC64;

Query Match 2.9%; Score 7; DB 16; Length 288;  
Best Local Similarity 100.0%; Pred. No. 1.4e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0

QY 194 RITPLS 200  
Db 201 RITPLS 207

RESULT 59  
R8P8A4 PRELIMINARY; PRT; 288 AA.

ID Q8P8A4  
AC Q8P8A4;  
DT 01-OCT-2002 (T-EMBLrel. 22, Created)  
DT 01-OCT-2002 (T-EMBLrel. 22, Last sequence update)  
DT 01-JUN-2003 (T-EMBLrel. 24, Last annotation update)  
DE Polyamine transport protein.  
GN POTH OR XCC2339.  
OS Xanthomonas campestris (pv. campestris).  
OC Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;  
OC Xanthomonadaceae; Xanthomonas.

```

OX NCBI_TaxID=340;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 33913 / NCPPB 528;
RX MEDLINE=22022145; PubMed=12024217;
RA da Silva A.C.R., Ferro J.A., Reinach F.C., Farah C.S., Furlan L.R.,
RA Quaggio R.B., Monteiro-Vitorello C.B., Van Sluys M.A., Almeida N.F.,
RA Alves L.M.C., de Amaral A.M., Bertolini M.C., Camargo L.E.A.,
RA Camarotte G., Cannavan F., Cardozo J., Chambergo F., Ciapina L.P.,
RA Cicarelli R.M.B., Coutinho L.L., Cursino-Santos J.R., El-Dorry H.,
RA Faria J.B., Ferreira A.J.S., Ferreira R.C.C., Ferro M.I.T.,
RA Formighieri E.F., Franco M.C., Greggio C.C., Gruber A.,
RA Katsuyama A.M., Kishi L.T., Leite R.P., Lemos E.C.M., Lemos M.V.F.,
RA Locali E.C., Machado M.A., Madeira A.M.B.N., Martinez-Rossi N.M.,
RA Martins E.C., Meidanis J., Menck C.F.M., Miyaki C.Y., Moon D.H.,
RA Moreira L.M., Novo M.T.M., Okura V.K., Oliveira M.C., Oliveira V.R.,
RA Pereira H.A., Rossi A., Sena J.A.D., Silva C., de Souza R.F.,
RA Spinola L.A.F., Takita M.A., Tamura R.E., Teixeira E.C., Tezza R.I.D.,
RA Trindade dos Santos M., Truffi D., Tsai S.M., White F.F.,
RA Secubal J.C., Kitajima J.P.;
RN "Comparison of the genomes of two Xanthomonas pathogens with differing
RT host specificities."
RL Nature 417:459-463(2002).
RL ENBL; AE012341; AAM41617.1; -.
DR GO; GO:0016020; C:membrane; IEA.
DR GO; GO:0005215; F:transporter activity; IEA.
DR GO; GO:0006810; P:transport; IEA.
DR InterPro; IPR000515; BPD_transp.
DR Pfam; PF00528; BPD_transp; 1.
DR PROSITE; PS00402; BPD_TRANSP_INN_MEMBER; 1.
RW Complete proteome.
SQ SEQUENCE 288 AA; 31877 MW; 1F37D3B6C71595A CRC64;

Query Match 2.9%; Score 7; DB 16; Length 288;
Best Local Similarity 100.0%; Pred. No. 1.4e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 194 RITLPLS 200
DB |||||
|||

RESULT 60
Q986W0 PRELIMINARY; PRT; 290 AA.
AC Q986W0;
DT 01-OCT-2001 (TRENBLrel. 18, Created)
DT 01-OCT-2001 (TRENBLrel. 18, Last sequence update)
DT 01-JUN-2003 (TRENBLrel. 24, Last annotation update)
DE ABC transporter, permease protein.
GN MLL7186.
OS Rhizobium loti (Mesorhizobium loti).
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC Phyllobacteriaceae; Mesorhizobium.
OX NCBI_TaxID=381;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MAFF303099;
RX MEDLINE=21082930; PubMed=11214968;
RA Kaneko T., Nakamura Y., Sato S., Asamizu E., Kato T., Sasamoto S.,
RA Watanabe A., Idegawa K., Ishikawa A., Kawashima K., Kimura T.,
RA Kishida Y., Kiyokawa C., Kohara M., Matsumoto M., Matsuno A.,
RA Mochizuki Y., Nakayama S., Nakazaki N., Shimo S., Sugimoto M.,
RA Takeuchi C., Yamada M., Tabata S.;
RT "Complete genome structure of the nitrogen-fixing symbiotic bacterium
RL Mesorhizobium loti."
RL Nature Res. 7:331-338(2000).
RL ENBL; AP003011; BAB53343.1; -.
DR GO; GO:0016020; C:membrane; IEA.
DR GO; GO:0005215; F:transporter activity; IEA.
DR GO; GO:0006810; P:transport; IEA.
DR InterPro; IPR000515; BPD_transp.
DR Pfam; PF00528; BPD_transp; 1.

```

```

DR PROSITE; PS00402; BPD_TRANSP_INN_MEMBER; 1.
KW Complete proteome.
SQ SEQUENCE 290 AA; 31559 MW; D231F51F73EDD7D2 CRC64;

Query Match 2.9%; Score 7; DB 16; Length 290;
Best Local Similarity 100.0%; Pred. No. 1.4e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 194 RITLPLS 200
DB |||||
|||

RESULT 61
Q97V08 PRELIMINARY; PRT; 291 AA.
AC Q97V08;
DT 01-OCT-2001 (TRENBLrel. 18, Created)
DT 01-OCT-2001 (TRENBLrel. 18, Last sequence update)
DT 01-JUN-2003 (TRENBLrel. 24, Last annotation update)
DE Hypothetical protein SSO2829.
GN SSO2829.
OS Sulfolobus solfataricus.
OC Archaea; Crenarchaeota; Thermoprotei; Sulfolobales; Sulfolobaceae;
OC Sulfolobus.
OX NCBI_TaxID=2287;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 35092 / DSM 1617 / P2;
RX MEDLINE=21332296; PubMed=1142726;
RA She Q., Singh R.K., Confalonieri F., Zivanovic Y., Allard G.,
RA Awayez M.J., Chan-Weiher C.C.-Y., Clausen I.G., Curtis B.A.,
RA De Moors A., Erauso G., Fletcher C., Gordon P.M.K.,
RA Heikamp-de Jong I., Jeffries A.C., Kozera C.J., Medina N., Peng X.,
RA Thi-Ngoc H.P., Redder P., Schenk M.E., Theriault C., Tolstrup N.,
RA Charlebois R.B., Doolittle W.F., Duguet M., Gaasterland T.,
RA Garrett R.A., Ragan M.A., Sensen C.W., Van der Oost J.;
RT "The complete genome of the crenarchaeon Sulfolobus solfataricus P2."
RL Proc. Natl. Acad. Sci. U.S.A. 98:7835-7840(2001).
DR ENBL; AE006877; AAK42939.1; -.
DR PIR; D90460; D90460.
DR GO; GO:0016020; C:membrane; IEA.
DR InterPro; IPR006685; MSion_channel.
DR Pfam; PF00924; MS_channel; 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 291 AA; 31912 MW; 2D1F8A1C3D42EA20 CRC64;

Query Match 2.9%; Score 7; DB 17; Length 291;
Best Local Similarity 100.0%; Pred. No. 1.4e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 158 VIIIAIG 164
DB |||||
|||

RESULT 62
Q9VVF5 PRELIMINARY; PRT; 302 AA.
AC Q9VVF5;
DT 01-MAY-2000 (TRENBLrel. 13, Created)
DT 01-MAY-2000 (TRENBLrel. 13, Last sequence update)
DT 01-MAY-2000 (TRENBLrel. 13, Last annotation update)
DE CG13272 protein.
GN CG13727.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Berkeley;

```

```

RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Ceiniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Ananides P.G., Scheer S.E., Li P.W., Hoskins R.A., Gallie R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner R.A., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Rendell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazey R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.D.G.,
RA April J.F., Agbayani A., An H.-J., Andrews-Pfannkuch C., Baldwin D.,
RA Ballev R.M., Basu A., Bakendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.V., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA Burtis K.C., Bueam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablo B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Foster C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jallali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Laslo P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusken D.R., Pacleb J.M., M.G.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese H.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Smith T.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svarskas R., Tecor C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Nassarman D.A., Weinstein G.M., Weissenbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang Q., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhou W., Zhou X., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster.";
DR EMBL; AB03524; AAF49356.1; -
DR FlyBase; FBgn003671; CG13727.
SQ SEQUENCE 302 AA; 31787 MW; 50373P9C24329B31 CRC64;

Query Match 2.9%; Score 7; DB 5; Length 302;
Best Local Similarity 100.0%; Pred. No. 1.5e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 64 SGEAGVP 70
DB 176 SGEAGVP 182

RESULT 63
Q23364 PRELIMINARY; PRT; 303 AA.
AC Q23364;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Hypothetical protein (Collagen).
GN ZC513.8 OR COL-43.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Pterodermidae; Caenorhabditis.
OX NCBI_TaxID=6239;
[1]
RN SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RX MEDLINE=99069613; PubMed=9851916;
RA None;
RT "Genome sequence of the nematode C. elegans: a platform for
investigating biology. The C. elegans Sequencing Consortium.";
RL Science 282:2012-2018(1998).

```

```

RN SEQUENCE FROM N.A.
RP STRAIN=Bristol N2;
RC Wu X., Le T.T.;
RT "The sequence of C. elegans cosmid ZC513.";
RL Submitted (APR-1996) to the EMBL/GenBank/DBJ databases.
[3]
RN SEQUENCE FROM N.A.
RP STRAIN=Bristol N2;
RC Waterston R.;
RA "Direct Submission.";
RT Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.
[4]
RN SEQUENCE FROM N.A.
RP Sando T., Ikeda T., Kagawa H.;
RT "HOMEOTIC PROTEINS, MAB-18 AND CBH-14 BI-DIRECTIONALLY REGULATE MALE-
TAIL COLLAGEN AND SPERM SPECIFIC PROTEIN GENES OF CAENORHABDITIS
ELEGANS.";
RT Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; U53155; AAC48270.1; -
DR EMBL; AB072926; BAB59889.1; -
DR F1R; T28999; T28999
DR WormPep; ZC513.8; CR07615.
DR InterPro; IPR008160; Collagen.
DR InterPro; IPR002486; Col_cuticle_N.
DR Pfam; PF01391; Collagen; 2.
DR Pfam; PF01484; Col_cuticle_N; 1.
KW Hypothetical protein.
SQ SEQUENCE 303 AA; 29454 MW; 4D036266F31097EB CRC64;

Query Match 2.9%; Score 7; DB 5; Length 303;
Best Local Similarity 100.0%; Pred. No. 1.5e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 235 GQAGTGP 241
DB 274 GQAGTGP 280

RESULT 64
Q89GJ0 PRELIMINARY; PRT; 313 AA.
AC Q89GJ0;
DT 01-JUN-2003 (TrEMBLrel. 24, Created)
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE ABC transporter permease protein.
GN BLI6355.
OS Bradyrhizobium japonicum.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC Bradyrhizobiaceae; Bradyrhizobium.
OX NCBI_TaxID=375;
[1]
RN SEQUENCE FROM N.A.
RC STRAIN=USDA 110;
RX MEDLINE=12597275;
RA Kaneko T., Nakamura Y., Sato S., Minamisawa K., Uchiumi T.,
RA Sasamoto S., Watanabe A., Idesawa K., Iriguchi M., Kawashima K.,
RA Kohara M., Matsumoto M., Shimo S., Teuruka H., Wada T., Yamada M.,
RA Tabata S.;
RT "Complete genomic sequence of nitrogen-fixing symbiotic bacterium
Bradyrhizobium japonicum USDA110.";
RL DNA Res. 9:189-197(2002).
DR EMBL; AP005958; BAC51620.1; -
DR GO; GO:0016020; C:membrane; IEA.
DR GO; GO:0005215; P:transporter activity; IEA.
DR GO; GO:0006810; P:transport; IEA.
DR InterPro; IPR000515; BPD_transp.
DR Pfam; PF00528; BPD_transp; 1.
DR PROSITE; PS00402; BPD_TRANS_INN_MEMBR; 1.
KW Complete proteome.
SQ SEQUENCE 313 AA; 34112 MW; AE97D4548EB8394A0 CRC64;

```

```

Query Match          2.9%; Score 7; DB 16; Length 313;
Best Local Similarity 100.0%; Pred. No. 1.6e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 193 QRITPL 199
DB 225 QRITPL 231

RESULT 65
O17035 PRELIMINARY; PRT; 314 AA.
AC O17035 (TREMBlrel. 05, Created)
DT 01-JAN-1998 (TREMBlrel. 05, Last sequence update)
DT 01-JAN-1998 (TREMBlrel. 05, Last sequence update)
DE 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
DE Hypothetical protein.
GN T15B7.5.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
[1]
RN SEQUENCE FROM N.A.
RP STRAIN=Bristol N2; PubMed=9851916;
RX MEDLINE=99069613;
RA None;
RT "Genome sequence of the nematode C. elegans: a platform for
investigating biology. The C. elegans Sequencing Consortium.";
RL Science 282:2012-2018(1998).
[2]
RN SEQUENCE FROM N.A.
RP STRAIN=Bristol N2;
RX MEDLINE=99069613;
RA None;
RT "The sequence of C. elegans cosmid T15B7.";
RL Submitted (SEP-1997) to the EMBL/GenBank/DBJ databases.
[3]
RN SEQUENCE FROM N.A.
RP STRAIN=Bristol N2;
RX MEDLINE=99069613;
RA None;
RT "Direct submission.";
RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
EMBL; AF022985; AAB69960.1; -.
DR PIR: T32247; T32247.
DR WormPep; T15B7.5; CR13656.
DR InterPro; IPR008160; Collagen.
DR InterPro; IPR002486; Col_cuticle_N.
DR Pfam; PF01391; Collagen; 3.
DR Pfam; PF01484; Col_cuticle_N; 1.
KW Hypothetical protein.
SQ SEQUENCE 314 AA; 30517 MW; 1CB91C0894314A7E CRC64;

Query Match          2.9%; Score 7; DB 5; Length 314;
Best Local Similarity 100.0%; Pred. No. 1.6e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 235 QOAGTPG 241
DB 251 QOAGTPG 257

RESULT 66
Q8VQ72 PRELIMINARY; PRT; 322 AA.
AC Q8VQ72;
DT 01-MAR-2002 (TREMBlrel. 20, Created)
DT 01-MAR-2002 (TREMBlrel. 20, Last sequence update)
DT 01-JUN-2003 (TREMBlrel. 24, Last annotation update)
DE Late competence protein ComGB.
GN COMGB.
OS Bacillus licheniformis.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxID=1402;

Query Match          2.9%; Score 7; DB 12; Length 323;
Best Local Similarity 100.0%; Pred. No. 1.6e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 31 AQSGLAS 37
DB 118 AQSGLAS 124

RESULT 68
Q8UIK1 PRELIMINARY; PRT; 326 AA.
AC Q8UIK1;
DT 01-JUN-2002 (TREMBlrel. 21, Created)
DT 01-JUN-2002 (TREMBlrel. 21, Last sequence update)
DT 01-JUN-2003 (TREMBlrel. 24, Last annotation update)
DE Iron-sulfur cluster binding protein.
GN ATU0292 OR AGR_C_502.
OS Agrobacterium tumefaciens (strain C58 / ATCC 33970).
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC Rhizobiaceae; Rhizobium/Agrobacterium group; Agrobacterium.
OX NCBI_TaxID=176299;
[1]
RN SEQUENCE FROM N.A.
RP STRAIN=ATCC14580;
RA Lapidus A.; Galleron N.; Andersen J.T.; Jorgensen P.L.; Ehrlich S.D.;
RA Sorokin A.;
RT "Co-linear scaffold of the Bacillus licheniformis and Bacillus
subtilis genomes and its use to compare their competence genes.";
RL FEMS Microbiol. Lett. 0:0-0(2002).
DR EMBL; AF459917; AAL67530.1; -.
DR GO; GO:0016020; C:membrane; IEA.
DR GO; GO:0015627; C:type II protein transporter activity; IEA.
DR GO; GO:0008565; F:protein transporter activity; IEA.
DR GO; GO:0015628; P:type II protein (Sec) secretion system; IEA.
DR InterPro; IPR001992; Bact_secr_systII.
DR InterPro; IPR003004; Bact_secr_systII.
DR Pfam; PF00482; GSPII_F; 1.
DR PRINTS; PR00812; BCTERIALGSPP.
SQ SEQUENCE 322 AA; 37096 MW; E9125C5D4C6283AE CRC64;

Query Match          2.9%; Score 7; DB 2; Length 323;
Best Local Similarity 100.0%; Pred. No. 1.6e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 NGHLYRE 14
DB 263 NGHLYRE 269

RESULT 67
Q9J1B0 PRELIMINARY; PRT; 323 AA.
AC Q9J1B0;
DT 01-OCT-2000 (TREMBlrel. 15, Created)
DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
DT 01-JUN-2003 (TREMBlrel. 24, Last annotation update)
DE Virus-cell attachment protein sigma C.
OS Nelson bay reovirus.
OC Viruses; dsRNA viruses; Reoviridae; Orthoreovirus.
OX NCBI_TaxID=118027;
[1]
RN SEQUENCE FROM N.A.
RP Duncan R.;
RT "Nelson bay reovirus S1 genome segment.";
RL Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.
EMBL; AF218360; AAF45159.1; -.
DR InterPro; IPR007662; Reo_sigmaC.
DR Pfam; PF04582; Reo_sigmaC; 1.
SQ SEQUENCE 323 AA; 34059 MW; 2DE9F780AE9AC358 CRC64;

Query Match          2.9%; Score 7; DB 12; Length 323;
Best Local Similarity 100.0%; Pred. No. 1.6e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 31 AQSGLAS 37
DB 118 AQSGLAS 124

RESULT 68
Q8UIK1 PRELIMINARY; PRT; 326 AA.
AC Q8UIK1;
DT 01-JUN-2002 (TREMBlrel. 21, Created)
DT 01-JUN-2002 (TREMBlrel. 21, Last sequence update)
DT 01-JUN-2003 (TREMBlrel. 24, Last annotation update)
DE Iron-sulfur cluster binding protein.
GN ATU0292 OR AGR_C_502.
OS Agrobacterium tumefaciens (strain C58 / ATCC 33970).
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC Rhizobiaceae; Rhizobium/Agrobacterium group; Agrobacterium.
OX NCBI_TaxID=176299;
[1]
RN SEQUENCE FROM N.A.
RP STRAIN=ATCC14580;
RA Lapidus A.; Galleron N.; Andersen J.T.; Jorgensen P.L.; Ehrlich S.D.;
RA Sorokin A.;
RT "Co-linear scaffold of the Bacillus licheniformis and Bacillus
subtilis genomes and its use to compare their competence genes.";
RL FEMS Microbiol. Lett. 0:0-0(2002).
DR EMBL; AF459917; AAL67530.1; -.
DR GO; GO:0016020; C:membrane; IEA.
DR GO; GO:0015627; C:type II protein transporter activity; IEA.
DR GO; GO:0008565; F:protein transporter activity; IEA.
DR GO; GO:0015628; P:type II protein (Sec) secretion system; IEA.
DR InterPro; IPR001992; Bact_secr_systII.
DR InterPro; IPR003004; Bact_secr_systII.
DR Pfam; PF00482; GSPII_F; 1.
DR PRINTS; PR00812; BCTERIALGSPP.
SQ SEQUENCE 322 AA; 37096 MW; E9125C5D4C6283AE CRC64;

```

```

RX MEDLINE=21608550; PubMed=11743193;
RA Wood D.W., Setubal J.C., Kaul R., Monks D.E., Kitajima J.P.,
RA Okura V.K., Zhou Y., Chen L., Wood G.E., Almeida N.F., Jr., Woo L.,
RA Chen Y., Paulsen I.T., Eisen J.A., Karp P.D., Bovee D. Sr.,
RA Chapman P., Clendenning J., Deatherage G., Gillet W., Grant C.,
RA Kutyavin T., Levy R., Li M.-J., McClelland E., Palmieri A., Gordon D.,
RA Raymond C., Rouse G., Saenphimmachak C., Wu Z., Romero P., Perry M.,
RA Zhang S., Yoo H., Tao Y., Biddle P., Jung M., Krespan W., Perry M.,
RA Gordon-Kamm B., Liao L., Kim S., Hendrick C., Zhao Z.-Y., Dolan M.,
RA Chumley F., Tingey S.V., Tomb J.-F., Gordon M.P., Olson M.V.,
RA Nester E.W.;
RT "The genome of the natural genetic engineer Agrobacterium tumefaciens
RT C58.";
RL Science 294:2317-2323 (2001).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=21608551; PubMed=11743194;
RA Goodner B., Hinkle G., Gattung S., Miller N., Blanchard M.,
RA Qurollo B., Goldman B.S., Cao Y., Askenazi M., Halling C., Mullin L.,
RA Houmlel K., Gordon J., Vaudin M., Iartchouk O., Epp A., Liu P.,
RA Wollam C., Allinger M., Doughty D., Scott C., Lappas C., Markelz B.,
RA Flanagan C., Crowell C., Gursion J., Lomo C., Sear C., Strub G.,
RA Cielo C., Slater S.;
RT "Genome sequence of the plant pathogen and biotechnology agent
RT Agrobacterium tumefaciens C58.";
RL Science 294:2323-2328 (2001).
DR EMBL; AE009001; AAL41314.1; -.
DR EMBL; AE007968; AAK86107.1; -.
DR PIR; AD2612; AD2612.
DR PIR; B97394; B97394.
DR GO; GO:0005489; F:electron transporter activity; IEA.
DR GO; GO:0006118; F:electron transport; IEA.
DR InterPro; IPR001450; 4Fe4S_ferredoxin.
DR Pfam; PF00037; fer4; 1.
DR PROSITE; PS00198; 4FE4S_FERREDOXIN; 1.
KW Complete proteome.
SQ SEQUENCE 326 AA; 36255 MW; 55514BB49B6FB4F7 CRC64;

Query Match 2.9%; Score 7; DB 16; Length 326;
Best Local Similarity 100.0%; Pred. No. 1.6e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 52 PDEDPRG 58
DB 33 PDEDPRG 39

RESULT 69
Q941W0 PRELIMINARY; PRT; 332 AA.
AC Q941W0;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE B1088C09.18 protein (P0446G04.1 protein).
GN B1088C09.18 OR P0446G04.1.
OS Oryza sativa (Rice), and
OS Oryza sativa (japonica cultivar-group).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Ehrhartoideae; Oryzaceae; Oryza.
OX NCBI_TaxID=4530, 39947;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Nipponbare;
RA Sasaki T., Matsumoto T., Yamamoto K.;
RT "Oryza sativa nipponbare (GA3) genomic DNA, chromosome 1, BAC
RL clone: B108C09.";
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Nipponbare;
RA Sasaki T., Matsumoto T., Yamamoto K.;

```

```

RT "Oryza sativa nipponbare (GA3) genomic DNA, chromosome 1, PAC
RT clone: P0446G04.";
RL Submitted (FEB-2001) to the EMBL/GenBank/DBSJ databases.
DR EMBL; AP003724; BAB68110.1; -.
DR EMBL; AP003452; BAB89578.1; -.
DR Gramene; Q941W0; -.
SQ SEQUENCE 332 AA; 38050 MW; 56400BF4409908A2 CRC64;

Query Match 2.9%; Score 7; DB 10; Length 332;
Best Local Similarity 100.0%; Pred. No. 1.6e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 119 RSEAAAV 125
DB 79 RSEAAAV 85

RESULT 70
Q8YWT5 PRELIMINARY; PRT; 332 AA.
AC Q8YWT5;
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Hypothetical protein All1509.
GN All1509.
OS Anabaena sp. (strain PCC 7120).
OC Bacteria; Cyanobacteria; Nostocales; Nostocaceae; Nostoc.
OX NCBI_TaxID=103650;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21595285; PubMed=11759840;
RA Kaneko T., Nakamura Y., Wolk C.P., Kuritz T., Sasamoto S.,
RA Watanabe A., Iriguchi M., Ishikawa A., Kawashima K., Kimura T.,
RA Kishida Y., Kohara M., Matsumoto M., Matsumoto A., Muraki A.,
RA Nakazaki N., Shimpo S., Sugimoto M., Takazawa M., Yamada M.,
RA Yasuda M., Tabata S.;
RT "Complete genomic sequence of the filamentous nitrogen-fixing
RT cyanobacterium Anabaena sp. strain PCC 7120.";
RL DNA Rep. 8:205-213 (2001).
DR EMBL; AP003586; BAB77875.1; -.
DR PIR; AH1994; AH1994.
DR InterPro; IPR005524; DUF318.
DR Pfam; PF03773; DUF318; 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 332 AA; 36033 MW; B4C6F3581AE16203 CRC64;

Query Match 2.9%; Score 7; DB 16; Length 332;
Best Local Similarity 100.0%; Pred. No. 1.6e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 161 IAIGAGI 167
DB 221 IAIGAGI 227

RESULT 71
Q7VNA0 PRELIMINARY; PRT; 336 AA.
AC Q7VNA0;
DT 01-OCT-2003 (TrEMBLrel. 25, Created)
DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Glyceral-3-phosphate dehydrogenase.
GN GPSA OR H00660.
OS Haemophilus ducreyi.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;
OC Pasteurellaceae; Haemophilus.
OX NCBI_TaxID=730;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=35000HP / ATCC 700724;
RA Munson R.S. Jr., Ray W.C., Mahairas G., Sabo P., Mungur R.,

```

RA Johnson L., Nguyen D., Wang J., Forst C., Hood L.;  
RT "The complete genome sequence of Haemophilus ducreyi";  
RL Submitted (JUN-2003) to the EMBL/GenBank/DBJ databases.  
RA EMBL; AE017152; AAP95585.1; -  
KW Complete proteome.  
SQ SEQUENCE 336 AA; 36575 MW; 70D4E2D2969126C4 CRC64;  
  
Query Match 2.9%; Score 7; DB 16; Length 336;  
Best Local Similarity 100.0%; Pred. No. 1.7e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 161 IAIGAGI 167  
DB 199 IAIGAGI 205  
  
RESULT 72  
O89PA5 PRELIMINARY; PRT; 338 AA.  
ID Q89PA5  
AC Q89PA5;  
DT 01-JUN-2003 (TrEMBLrel. 24, Created)  
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)  
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
DE BL3577 protein.  
GN BLR3577  
OS Bradyrhizobium japonicum.  
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;  
OC Bradyrhizobiaceae; Bradyrhizobium.  
OX NCBI\_TaxID=375;  
[1]  
RN SEQUENCE FROM N.A.  
RP STRAIN=USDA 110;  
RC MEDLINE=22484998; PubMed12597275;  
RA Kaneko T., Nakamura Y., Sato S., Minamisawa K., Uchiumi T.,  
RA Sasamoto S., Watanabe A., Idesawa K., Iriiguchi M., Kawashima K.,  
RA Kohara M., Matsumoto M., Shimpō S., Tsuruoka H., Wada T., Yamada M.,  
RA Tabata S.;  
RT "Complete genomic sequence of nitrogen-fixing symbiotic bacterium  
Bradyrhizobium japonicum USDA110";  
RL DNA Res. 9:189-197(2002).  
DR EMBL; AP005948; BAC48842.1; -  
DR GO; GO:0004371; F:glycerol kinase activity; IEA.  
DR GO; GO:0006071; F:glycerol metabolism; IEA.  
DR InterPro; IPR004006; Dakt.  
DR Pfam; PF02733; Dakt1; 1.  
KW Complete proteome.  
SQ SEQUENCE 338 AA; 35577 MW; CF09F6CF6AB97576 CRC64;  
  
Query Match 2.9%; Score 7; DB 16; Length 338;  
Best Local Similarity 100.0%; Pred. No. 1.7e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 63 VSGAGV 69  
DB 104 VSGAGV 110  
  
RESULT 73  
Q8YTX8 PRELIMINARY; PRT; 346 AA.  
ID Q8YTX8  
AC Q8YTX8;  
DT 01-MAR-2002 (TrEMBLrel. 20, Created)  
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)  
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
DE Iron(III) dicitrate transport system permease protein.  
GN ALI2585.  
OS Anabaena sp. (strain PCC 7120).  
OC Bacteria; Cyanobacteria; Nostocales; Nostocaceae; Nostoc.  
OX NCBI\_TaxID=103690;  
[1]  
RN SEQUENCE FROM N.A.  
RP MEDLINE=21595285; PubMed=11759840;  
RA Kaneko T., Nakamura Y., Woik C.P., Kuritz T., Sasamoto S.,

Watanabe A., Iriiguchi M., Iahikawa A., Kawashima K., Kimura T.,  
Kishida Y., Kohara M., Matsumoto M., Matsuno A., Muraki A.,  
RA Nakazaki N., Shimpō S., Sugimoto M., Takazawa M., Yamada M.,  
RA Yasuda M., Tabata S.;  
RT "Complete genomic sequence of the filamentous nitrogen-fixing  
cyanobacterium Anabaena sp. strain PCC 7120";  
RL DNA Res. 8:205-213(2001).  
DR EMBL; AP003590; BA574284.1; -  
DR F1R; AB2129; AS2129.  
DR GO; GO:0016020; C:membrane; IEA.  
DR GO; GO:0005215; P:transporter activity; IEA.  
DR GO; GO:0006810; P:transport; IEA.  
DR InterPro; IPR000522; FecD.  
DR Pfam; PF01032; FecCD; 1.  
DR ProDom; PD001557; FecCD; 1.  
KW Complete proteome.  
SQ SEQUENCE 346 AA; 36510 MW; ED39A2EDDC42D0F3D CRC64;  
  
Query Match 2.9%; Score 7; DB 16; Length 346;  
Best Local Similarity 100.0%; Pred. No. 1.7e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 196 TLPLSAP 202  
DB 133 TLPLSAP 139  
  
RESULT 74  
Q8KY10 PRELIMINARY; PRT; 351 AA.  
ID Q8KY10  
AC Q8KY10;  
DT 01-OCT-2002 (TrEMBLrel. 22, Created)  
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)  
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
DE RepB.  
GN REPB.  
OS Rhizobium etli.  
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;  
OC Rhizobiaceae; Rhizobium/Agrobacterium group; Rhizobium.  
OX NCBI\_TaxID=29449;  
[1]  
RN SEQUENCE FROM N.A.  
RP STRAIN=CE3;  
RA Cevallos M.A., Izquierdo J., Porta H., Tun C., Davila G., Brom S.;  
RT "Rhizobium etli CE3 contains at least three plasmids of the RepABC  
family: A structural and an evolutionary analysis";  
RL Submitted (OCT-2000) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AF313446; AAM88942.1; -  
DR GO; GO:0005727; C:extrachromosomal circular DNA; IEA.  
DR GO; GO:0003677; F:DNA binding; IEA.  
DR GO; GO:0030542; P:plasmid partitioning (sensu Bacteria); IEA.  
DR InterPro; IPR003115; ParBc.  
DR InterPro; IPR004437; ParB\_part.  
DR Pfam; PF02195; ParBc; 1.  
DR SMART; SM00470; ParB; 1.  
DR TIGRFAMs; TIGR00180; parB\_part; 1.  
KW Plasmid.  
SQ SEQUENCE 351 AA; 38860 MW; 041CB303304830EF CRC64;  
  
Query Match 2.9%; Score 7; DB 2; Length 351;  
Best Local Similarity 100.0%; Pred. No. 1.7e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 122 AAAPQPV 128  
DB 284 AAAPQPV 290  
  
RESULT 75  
Q9GUI1 PRELIMINARY; PRT; 355 AA.  
ID Q9GUI1  
AC Q9GUI1;



DT 01-MAR-2001 (TrEMBLrel. 16, Created)  
 DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)  
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)  
 DE Y55DSA.3 protein.  
 GN Y55DSA.3.  
 OS Caenorhabditis elegans.  
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;  
 OC Rhabditidae; Peloderinae; Caenorhabditis.  
 OX NCBI\_TaxID=6239;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Bristol N2;  
 RX MEDLINE=99069613; PubMed=9851916;  
 RA None;  
 RT "Genome sequence of the nematode C. elegans: a platform for  
 RT investigating biology. The C. elegans Sequencing Consortium.";  
 RL Science 282:2012-2018(1998).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Bristol N2;  
 RA Waterston R.;  
 RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AC084196; AAK39622.1; -.  
 DR WormPep; Y55DSA.3; CE27498.  
 DR InterPro; IPR003199; Chigly\_hydrolase.  
 DR Pfam; PF02275; CBAH; 1.  
 SQ SEQUENCE 355 AA; 40567 MW; FAB51D5F49393A73 CRC64;

Query Match 2.9%; Score 7; DB 5; Length 355;  
 Best Local Similarity 100.0%; Pred. No. 1.7e+02;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 91 FTTEIQE 97  
 Db 78 FTTEIQE 84

Search completed: March 17, 2004, 07:26:18  
 Job time : 51 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 17, 2004, 07:08:04 ; Search time 60 Seconds

(without alignments)  
1139.608 Million cell updates/sec

Title: US-10-057-951-2\_COPY\_22\_263

Perfect score: 242

Sequence: 1 SGGFWNGHLYREDQSPA.....PVPQSGSLMQAGTPGA 242

Scoring table: OLIGO Gapop 60.0 , Gapext 60.0

Searched: 1586107 seqs, 282547505 residues

Word size : 0

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 100 summaries

Database : A\_Geneseq\_23Jan04.\*

1: Geneseqp1980s.\*

2: Geneseqp1980s.\*

3: Geneseqp2000s.\*

4: Geneseqp2001s.\*

5: Geneseqp2002s.\*

6: Geneseqp2003as.\*

7: Geneseqp2003bs.\*

8: Geneseqp2004s.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	242	100.0	263	2	AAW87769 Human tis
2	242	100.0	263	2	AAW87769 Human tis
3	242	100.0	263	4	AAE00300 Human tis
4	242	100.0	263	5	ABR40414 Human sec
5	229	94.6	263	5	AAU86149 Human PRO
6	172	71.1	263	3	ABR43237 Human ORF
7	110	45.5	146	5	ABR40487 Human sec
8	110	45.5	146	5	ABR40561 Human sec
9	101	41.7	263	4	AAW93748 Human pol
10	66	27.3	66	4	AAW18800 Peptide #
11	66	27.3	66	4	ABR37905 Peptide #
12	66	27.3	66	4	AAW13114 Peptide #
13	66	27.3	66	4	ABR23159 Protein #
14	66	27.3	66	4	AAW71037 Human bon
15	66	27.3	66	4	AAW58537 Human bra
16	66	27.3	66	4	ABG52752 Human liv
17	66	27.3	66	5	ABG40828 Human pep
18	47	19.4	286	2	AAW05220 Kringle1
19	34	14.0	55	2	AAW12397 Human 5'
20	32	13.2	56	2	AAW12615 Human 5'
21	20	8.3	39	2	AAW72640 Nervous g
22	20	8.3	39	2	AAW72641 Nervous q
23	13	5.4	81	7	ABR42624 Human kri
24	10	4.1	527	2	AAW54157 t-PA muta
25	10	4.1	527	2	AAW54154 t-PA muta

Abu16473	Protein e
Abm72983	Staphyloc
Abm69063	Phototrab
Abu23579	Protein e
Abu56600	Mouse L5/
Abu14271	Mouse gro
Abu66601	Mouse L5/
Abu14272	Mouse gro
Abu31156	Murine ma
Abu82790	Mouse MSP
Abu34603	Protein e
Abu36546	Protein e
Abu34316	Protein e
Abu32528	West Nile
Abu1772	Peptide #
Abu35571	Peptide #
Abu75460	Human bon
Abu62644	Human bra
Abu57210	Human liv
Abu45027	Human pep
Abu9077	AP-2gamma
Abu5213	FtsZ frag
Abu46725	Propionib
Abu43244	Propionib
Abu7491	Human ORF
Abu04700	Novel hum
Abu31178	Protein e
Abu70881	C. Glutam
Abu90034	C glutam
Abu32517	Novel hum
Abu35428	Chlamydia
Abu97730	Human ORF
Abu96351	E. faeciu
Abu35431	Chlamydia
Abu17162	Novel sig
Abu84513	Human imm
Abu87649	Novel cen
Abu37373	Novel cen
Abu33870	Human nov
Abu70309	Phototrab
Abu4184	Staphyloc
Abu41811	Human pol
Abu41810	Human pol
Abu59873	Drosophil
Abu16506	Protein e
Abu17587	Novel hum
Abu36580	Staphyloc
Abu1753	Staphyloc
Abu68531	Drosophil
Abu06112	Nelson Ba
Abu41611	Human ORF
Abu40024	Human pol
Abu35472	Haemophil
Abu30312	Protein e
Abu68792	Phototrab
Abu5426	E. faeciu
Abu30030	Protein e
Abu90557	Chlamydia
Abu94377	E. faeciu
Abu37309	E. faeciu
Abu01333	FtsZ1 pol
Abu56062	Propionib
Abu3405	Novel hum
Abu16425	Novel hum
Abu45181	Protein e
Abu20208	Protein e
Abu47610	Protein e
Abu38368	Salmonell

99 7 2.9 465 6 ABU28051 Protein e  
100 7 2.9 469 6 ABU32175 Protein e

ALIGNMENTS

RESULT 1  
AAW87769 ID AAW87769 standard; protein; 263 AA.  
AC AAW87769;  
XX  
DT 29-MAR-1999 (first entry)  
XX  
DE Human tissue plasminogen activator-like protease t-PALP.  
XX  
KW Tissue plasminogen activator-like protease; t-PALP; human;  
KW circulatory system-related disorder; blood clotting; stroke; thrombosis;  
KW peripheral arterial occlusion; pulmonary embolism; myocardiothrombosis;  
KW diagnosis; therapy.  
XX  
OS Homo sapiens.

XX Key Location/Qualifiers  
FH Peptide 1..21  
FT /label= Sig\_peptide  
FT Protein 22..263  
FT /label= Mat\_protein  
FT Peptide 22..31  
FT /note= "epitope-bearing region"  
FT Domain 25..84  
FT /note= "kringle domain"  
FT Peptide 35..44  
FT /note= "epitope-bearing region"  
FT Peptide 71..81  
FT /note= "epitope-bearing region"  
FT Domain 85..263  
FT /note= "protease domain"  
FT Peptide 91..107  
FT /note= "epitope-bearing region"  
FT Peptide 119..128  
FT /note= "epitope-bearing region"  
FT Peptide 138..147  
FT /note= "epitope-bearing region"  
FT Peptide 155..167  
FT /note= "epitope-bearing region"  
FT Peptide 193..203  
FT /note= "epitope-bearing region"  
FT Peptide 206..215  
FT /note= "epitope-bearing region"  
FT Peptide 227..237  
FT /note= "epitope-bearing region"  
FT Peptide 243..252  
FT /note= "epitope-bearing region"

XX W09854199-A1.  
PN  
XX  
PD 03-DEC-1998.  
XX  
XX 27-MAY-1998; 98WO-US010728.  
XX  
XX 28-MAY-1997; 97US-0048000P.  
XX  
XX (HUMA-) HUMAN GENOME SCI INC.  
XX  
XX Ebner R, Moore PA, Ruben SM;  
PI WPI; 1999-070207/06.  
XX N-PSDB; AAV99636.  
DR  
DR  
XX  
XX New tissue plasminogen activator-like protease - useful in the diagnosis  
PT and treatment of circulatory system-related disorders.

XX Claim 1; Page 56-57; 76pp; English.  
XX  
CC This is the amino acid sequence of tissue plasminogen activator-like  
CC protease (t-PALP), a novel member of the serine protease family that  
CC shares sequence homology to human tissue plasminogen activator (see  
CC AAW87770). The t-PALP sequence was deduced from a cDNA clone (see  
CC AAV99636) derived from activated monocytes. The 2.5 kb t-PALP message has  
CC also been detected in heart, brain, lung, placenta, liver, skeletal  
CC muscle, kidney, pancreas, spleen, thymus, prostate, testis, ovary, small  
CC intestine, colon and peripheral blood leukocytes. Isolated nucleic acids  
CC encoding amino acids 21 to 242, -20 to 242, 1-242, 4-63 (kringle domain)  
CC and 64-242 (protease domain) of t-PALP, or encoding epitope-bearing  
CC portions of t-PALP, are also claimed, as are recombinant vectors, host  
CC cells, and methods for producing t-PALP polypeptides. t-PALP may be used  
CC to detect and treat disorders related to the circulatory system, and to  
CC identify agonists and antagonists of t-PALP activity. The homology  
CC between t-PALP and tPA indicates that t-PALP may be involved in the  
CC regulation of normal and abnormal clotting in e.g. stroke, deep-vein  
CC thrombosis, peripheral arterial occlusion, pulmonary embolism and  
CC myocardiothrombosis  
XX  
SQ Sequence 263 AA;

Query Match 100.0%; Score 242; DB 2; Length 263;  
Best Local Similarity 100.0%; Pred. No. 3.4e-208;  
Matches 242; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 SGGCFWNGHLYREDOTSPAGLRLCLNWLDAQSLASAPVSGAGNHSYCRNPDEDPGFW 60  
DB 22 SGGCFWNGHLYREDOTSPAGLRLCLNWLDAQSLASAPVSGAGNHSYCRNPDEDPGFW 81  
QY 61 CYVSGEAGVPEKPCEDLRCPTTSQALPAFTTEIQEASEGPGADEVQVAFANALPARS 120  
DB 82 CYVSGEAGVPEKPCEDLRCPTTSQALPAFTTEIQEASEGPGADEVQVAFANALPARS 141  
QY 121 EAAAVQVIGISORVRMNSKEKDLGLGVLTGVTWVILIAIGAGILGYSYKGGDLK 180  
DB 142 EAAAVQVIGISORVRMNSKEKDLGLGVLTGVTWVILIAIGAGILGYSYKGGDLK 201  
QY 181 EQHDQKVCEREMORITLPLSAFTNPTCEIVDEKTVVHTSQTVPDQEGSTPLMGQAGTP 240  
DB 202 EQHDQKVCEREMORITLPLSAFTNPTCEIVDEKTVVHTSQTVPDQEGSTPLMGQAGTP 261  
QY 241 GA 242  
DB 262 GA 263

RESULT 2  
AAV05219 ID AAV05219 standard; protein; 263 AA.  
XX AC AAV05219;  
XX DT 17-JUN-1999 (first entry)  
XX DE Kringle1 protein sequence.  
XX  
KW Kringle1; cancer; inflammation; autoimmunity; rheumatoid arthritis; AIDS;  
KW CNS inflammation; cerebellar degeneration; Alzheimer's disease; asthma;  
KW Parkinson's disease; multiple sclerosis; amyotrophic lateral sclerosis;  
KW neurological abnormality; ischaemia reperfusion injury; ischaemic injury;  
KW cardiovascular disease; kidney disease; liver disease; aplastic anaemia;  
KW myocardial infarction; hypertension; hyperextension; allergy; infection;  
KW myelodysplastic syndrome; haematologic abnormality; diagnosis; therapy;  
XX male pattern baldness.  
OS Homo sapiens.  
PN W09911788-A1.  
XX  
XX 11-MAR-1999.  
PD

```

XX 02-SEP-1998; 98WO-US018270.
PF 02-SEP-1997; 97US-0056032P.
PR 01-SEP-1998; 98US-00144889.
XX (SMIK) SMITHLINE BEECHAM CORP.
XX Albone EF, Kikly KK;
XX WPI; 1999-214707/18.
DR N-PSDB; AAX28354.
XX New kringle1 polypeptides and polynucleotides.
XX Claim 1; Page 31-32; 42pp; English.
XX This sequence is a Kringle1 polypeptide of the invention. The kringle1
XX polypeptides (I) are used to screen for agonists and antagonists.
XX Agonists are used to treat subjects in need of enhanced activity or
XX expression of (I). Antagonists are used to treat subjects having need to
XX inhibit the activity or expression of (I). The methods can be used to
XX treat conditions such as cancer, inflammation, autoimmunity, allergy,
XX asthma, rheumatoid arthritis, CNS inflammation, cerebellar degeneration,
XX Alzheimer's disease, Parkinson's disease, multiple sclerosis,
XX amyotrophic lateral sclerosis, head injury damage and other neurological
XX abnormalities, ischemia reperfusion injury, cardiovascular disease,
XX kidney disease, liver disease, ischemic injury, myocardial infarction,
XX hypotension, hypertension, AIDS, myelodysplastic syndromes and other
XX hematologic abnormalities, aplastic anemia, male pattern baldness, and
XX bacterial, fungal, protozoan and viral infections. The kringle1
XX polypeptides may also be used to generate antibodies. Determining the
XX presence or absence of mutations in, and analysing for the presence or
XX absence of expression of, kringle1 polynucleotides can be used to
XX diagnose a disease or susceptibility to a disease related to expression
XX or activity of kringle1 proteins. The polynucleotides may also be used
XX for chromosome identification, and mapping
XX
XX Sequence 263 AA;
XX
Query Match 100.0%; Score 242; DB 2; Length 263;
Best Local Similarity 100.0%; Pred. No. 3 4e-208;
Matches 242; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 SGGCFWNGHLYREDQTSFAPGLRCLNWLDAQSGLASAPVSGAGNHSYCNPNPDRGPW 60
DB 22 SGGCFWNGHLYREDQTSFAPGLRCLNWLDAQSGLASAPVSGAGNHSYCNPNPDRGPW 81
QY 61 CVVSGEAGVPEKPCEDLRCPTTSQALPAFTTEIQASGPGADEVQVFAPANALPARS 120
DB 82 CVVSGEAGVPEKPCEDLRCPTTSQALPAFTTEIQASGPGADEVQVFAPANALPARS 141
QY 121 EAAVQPVIGISQVRMNSKKDLGTLGYVLGTTMNVIIAIGAGIILGYSYKRGKDLK 180
DB 142 EAAVQPVIGISQVRMNSKKDLGTLGYVLGTTMNVIIAIGAGIILGYSYKRGKDLK 201
QY 181 EOHDKVCEREMQRTILPLSAFTNPTCEIIVDEKTVVHTSQTVPDQESGTPLMGQAGTF 240
DB 202 EOHDKVCEREMQRTILPLSAFTNPTCEIIVDEKTVVHTSQTVPDQESGTPLMGQAGTF 261
QY 241 GA 242
DB 262 GA 263
XX
RESULT 3
AAE00300
ID AAE00300 standard; protein; 263 AA.
XX
AC AAE00300;
XX
DT 13-JUN-2001 (first entry)
XX

```

```

DS Human tissue-plasminogen activator-like protease (t-PALP).
XX
KW Human; tissue-plasminogen activator-like protease; t-PALP; therapy;
KW vascular disease; stroke; deep vein thrombosis; keloid; asthma;
KW arterial occlusion; blood coagulation disorder; cerebroprotective;
KW autoimmune system disorder; human immunodeficiency syndrome; cytostatic;
KW rheumatoid arthritis; graft-versus-host disease; thyroiditis; cardiant;
KW insulin dependent diabetes; inflammatory eye disease; allergy; antiviral;
KW cardiovascular disease; heart disease; arrhythmia; myocardial ischaemia;
KW hyperproliferative disorder; hypertrophic scar; neurological disease;
KW Creutzfeldt-Jakob syndrome; neurodegenerative disorder; antiinflammatory;
KW Alzheimer's disease; Parkinson's disease; immunosuppressive; antifungal;
KW infectious disease; drug screening; gene therapy; neuroprotective;
XX cancer; ophthalmological; antibacterial; vulnery.
XX Homo sapiens.
XX
Key Location/Qualifiers
Binding-site 1..165 /note= "Binds to FLAG polypeptide to form t-PALP-FLAG
fusion protein"
Peptide 1..21 /label= "Signal_peptide"
Domain 4..63 /label= "Kingle_domain"
Region 12..21 /note= "Conserved region"
Protein 22..263 /note= "Human mature tissue-plasminogen activator-like
protease (t-PALP); Binds to FLAG polypeptide to form t-
PALP-FLAG fusion protein"
Region 22..38 /note= "Conserved region"
Region 22..31 /note= "Epitope-bearing portion"
Region 35..44 /note= "Epitope-bearing portion"
Region 39..49 /note= "Conserved region"
Region 50..62 /note= "Conserved region"
Region 63..84 /note= "Conserved region"
Domain 64..242 /label= "Protease_domain"
Region 71..81 /note= "Epitope-bearing portion"
Region 85..97 /note= "Conserved region"
Region 91..107 /note= "Epitope-bearing portion"
Region 100..118 /note= "Conserved region"
Region 119..128 /note= "Epitope-bearing portion"
Region 119..127 /note= "Conserved region"
Region 128..143 /note= "Conserved region"
Region 138..147 /note= "Epitope-bearing portion"
Region 146..163 /note= "Conserved region"
Region 155..167 /note= "Epitope-bearing portion"
Region 164..180 /note= "Conserved region"
Region 186..200 /note= "Conserved region"
Region 193..203 /note= "Epitope-bearing portion"
Region 201..220 /note= "Conserved region"
Region 201..220 /note= "Conserved region"

```

PT Region 206...215  
 FT /note= "Epitope-bearing portion"  
 FT 221...236  
 FT /note= "Conserved region"  
 FT 227...237  
 FT /note= "Epitope-bearing portion"  
 FT 237...248  
 FT /note= "Conserved region"  
 FT 243...252  
 FT /note= "Epitope-bearing portion"  
 FT 249...263  
 FT /note= "Conserved region"  
 XX  
 XX WO200125252-A1.  
 XX  
 XX PD 12-APR-2001.  
 XX  
 XX PF 03-OCT-2000; 2000WO-US027239.  
 XX  
 XX PR 04-OCT-1999; 99US-00411977.  
 XX  
 XX PA (HUMA-) HUMAN GENOME SCI INC.  
 XX  
 XX PI Moore PA, Ruben SM, Ebner R;  
 XX  
 XX DR WPI; 2001-235402/24.  
 XX  
 XX DR N-PSDB; AAD03460.  
 XX  
 XX New (gene encoding and antibody immunospecific for a) tissue-plasminogen  
 PT activator-like protease, useful for the diagnosis and treatment of  
 PT (cardio)vascular diseases, hyperproliferative disorders, immune system  
 PT disorders and cancers.  
 XX  
 XX PS Claim 17; Fig 1; 323pp; English.  
 XX  
 XX The present amino acid sequence is HMSIB42 clone human tissue-plasminogen  
 CC activator-like protease (t-PALP). The t-PALP sequence and their  
 CC (ant)agonists are useful for the diagnosis and treatment of vascular  
 CC diseases e.g. stroke, deep vein thrombosis and arterial occlusion, blood  
 CC coagulation disorders, (auto)immune system disorders e.g. human  
 CC immunodeficiency syndrome, rheumatoid arthritis, graft-versus-host  
 CC disease, thyroiditis, insulin dependent diabetes and inflammatory eye  
 CC disease, allergic reactions e.g. asthma, cardiovascular diseases e.g.  
 CC heart disease, arrhythmia and myocardial ischemia, hyperproliferative  
 CC disorders, cancers, hypertrophic scars and keloids, neurological diseases  
 CC e.g. Creutzfeldt-Jakob syndrome, neurodegenerative disorders e.g.  
 CC Alzheimer's disease and Parkinson's disease and infectious disease e.g.  
 CC viral, bacterial and fungal infections. The t-PALP sequences are also  
 CC useful for drug screening. The t-PALP nucleotides are useful as  
 CC chromosome markers and are involved in gene therapy  
 XX  
 XX SQ Sequence 263 AA;  
 Query Match 100.0%; Score 242; DB 4; Length 263;  
 Best Local Similarity 100.0%; Pred. No. 3.4e-208;  
 Matches 242; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 SGGCFWNGHLYREDOTSPAGLCLNWLDAQSLAPVSGAGNHSYCRNPDPGRGW 60  
 DB 22 SGGCFWNGHLYREDOTSPAGLCLNWLDAQSLAPVSGAGNHSYCRNPDPGRGW 81  
 QY 61 CVVSGEAGVPEKPCEDLRCPTTSQALPAFTTBIQESBPGGADEVQVFPANALPARS 120  
 DB 82 CVVSGEAGVPEKPCEDLRCPTTSQALPAFTTBIQESBPGGADEVQVFPANALPARS 141  
 QY 121 EAAAVQVIGISQVRMNSKKDLGTGYVLGTMMVIIAIGAGIILGYSKRGDLK 180  
 DB 142 EAAAVQVIGISQVRMNSKKDLGTGYVLGTMMVIIAIGAGIILGYSKRGDLK 201  
 QY 181 EQHDKQVCERENQRTILPLSAFTNPTCEIVDEKTVVHTSQTVPDQEGSTPLMGAGTP 240  
 DB 202 EQHDKQVCERENQRTILPLSAFTNPTCEIVDEKTVVHTSQTVPDQEGSTPLMGAGTP 261

QY 241 GA 242  
 DB 262 GA 263  
 RESULT 4  
 ABR40414  
 ID ABR40414 standard; protein; 263 AA.  
 XX AC ABR40414;  
 XX 13-JUN-2003 (first entry)  
 XX Human secreted protein #SEQ ID 164.  
 XX  
 XX Human; secreted protein; anti-HIV; nootropic; neuroprotective;  
 KW anti-anginal; immunosuppressive; immunomodulator; cytostatic; cardiant;  
 KW hepatotropic; antiinflammatory; antiallergic; antidiabetic;  
 KW gastrointestinal; antifertility; nephrotropic; virucide; hypotensive;  
 KW vasotropic; dermatological; osteopathic; antiarthritic; antiparkinsonian;  
 KW antiasthmatic; antipsoriatic; cererebroprotective; antibacterial;  
 KW fungicide; antiparasitic; protein therapy; gene therapy; cancer;  
 KW hyperproliferative disorder; leukaemia; autoimmune disorder;  
 KW immunodeficiency; HIV; infertility; uterine disorders; endometriosis;  
 KW amenorrhea; ocular disorder; neurological disorder; wound healing;  
 KW Huntington's disease; gastrointestinal disorder; inflammatory disease;  
 KW Crohn's disease; kidney disorder; renal failure; diabetes; obesity;  
 KW cirrhosis; hepatitis; cardiovascular disorder; angina; osteoarthritis;  
 KW rheumatoid arthritis; psoriasis; multiple sclerosis; asthma;  
 KW respiratory disorder; infectious disease; chromosome identification;  
 XX food additive; nutrition.  
 XX  
 XX OS Homo sapiens.  
 XX  
 XX WO200268628-A1.  
 XX  
 XX 06-SEP-2002.  
 XX  
 XX 21-FEB-2002; 2002WO-US005301.  
 XX  
 XX 23-FEB-2001; 2001US-0270625P.  
 XX 12-JUN-2001; 2001US-030417P.  
 XX  
 XX (HUMA-) HUMAN GENOME SCI INC.  
 XX  
 XX Rosen CA, Komatsoulis G, Baker KP, Fiscella M, Moore PA, Wei P;  
 XX Duan DR, Shi Y, Gupta R;  
 XX WPI; 2002-750417/81.  
 XX N-PSDB; ABZ82469.  
 XX  
 XX New human secreted proteins and nucleic acids, useful for preventing,  
 PT treating, or ameliorating a medical condition, e.g. cancer, autoimmune  
 PT disorder, HIV, Parkinson's disease, Alzheimer's disease, diabetes,  
 PT obesity or cirrhosis.  
 XX  
 XX Claim 11; Page 755; 873pp; English.  
 XX  
 XX The invention relates to novel human secreted proteins and the genes  
 CC encoding them. Genes and proteins of the invention may be useful for  
 CC preventing, treating or ameliorating medical conditions e.g. by protein  
 CC or gene therapy. These conditions include cancer and hyperproliferative  
 CC disorders, immune cell proliferative disorders (e.g. leukaemia),  
 CC autoimmune disorders, immunodeficiencies (e.g. HIV infection),  
 CC infertility, placental and uterine disorders (e.g. endometriosis),  
 CC amenorrhea, ocular disorders, neurological disorders (e.g. Huntington's  
 CC disease), wound healing, gastrointestinal system disorders, particularly  
 CC inflammatory diseases (e.g. Crohn's disease), kidney disorders (e.g.  
 CC renal failure), diabetes, obesity, cirrhosis, hepatitis, cardiovascular  
 CC disorders (e.g. angina), rheumatoid arthritis, osteoarthritis, psoriasis,  
 CC multiple sclerosis, respiratory disorders (e.g. asthma), or infectious  
 CC diseases caused by bacterial, parasitic, viral or fungal agents. The  
 CC nucleic acids are also useful for chromosome identification, radiation

CC hybrid mapping or long-range restriction mapping. The polypeptide,  
CC polynucleotide, agonist or antagonist may be used as a food additive  
CC or preservative to increase or decrease storage capabilities, fat content  
CC or other nutritional components. The sequences given in records ABR40409-  
CC ABR40590 and ABR282464-ABR282611 represent human secreted proteins and the  
CC genes encoding them  
XX  
SQ Sequence 263 AA;  
  
Query Match 100.0%; Score 242; DB 5; Length 263;  
Best Local Similarity 100.0%; Pred. No. 3.4e-208;  
Matches 242; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 1 SGGCFWNGHLYREDQTSAPGLRCLNWLDAQSLASAPVSGAGNHSYCRNPDEPRGW 60  
Db 22 SGGCFWNGHLYREDQTSAPGLRCLNWLDAQSLASAPVSGAGNHSYCRNPDEPRGW 61  
  
QY 61 CYVSGEAGVPEKPCEDLRCPETTSQALPAFTTEIQEASGPGADVQVAFANALPARS 120  
Db 82 CYVSGEAGVPEKPCEDLRCPETTSQALPAFTTEIQEASGPGADVQVAFANALPARS 141  
  
QY 121 EAAAVQPVIGISQVRVNSKEKDLGLTGLVGLITMVMVIIAIGAGIILGYSYKRGKDLK 180  
Db 142 EAAAVQPVIGISQVRVNSKEKDLGLTGLVGLITMVMVIIAIGAGIILGYSYKRGKDLK 201  
  
QY 181 EQHDQKVCEREMQRTILPLSAFTNPTCEIVDEKTVVHTSQTPVDPQEGSTPLMGQATP 240  
Db 202 EQHDQKVCEREMQRTILPLSAFTNPTCEIVDEKTVVHTSQTPVDPQEGSTPLMGQATP 261  
  
QY 241 GA 242  
Db 262 GA 263  
  
RESULT 5  
AAU86149  
ID AAU86149 standard; protein; 263 AA.  
AC AAU86149;  
XX  
XX 15-JUL-2002 (first entry)  
DE Human PRO264 polypeptide.  
XX  
XX Human; PRO; benign tumour; malignant tumour; lymphoid malignancy;  
KW leukaemia; neuronal disorder; stromal disorder; blastocoelec disorder;  
KW inflammatory disorder; immune disorder; angiogenic disorder; cytostatic;  
KW neuroprotective.  
OS Homo sapiens.  
XX  
XX WC200153486-A1.  
XX  
XX 26-JUL-2001.  
XX  
XX 11-FEB-2000; 2000WO-US003565.  
XX  
XX 08-MAR-1999; 99WO-US005028.  
PR 11-MAR-1999; 99US-0123972P.  
PR 11-MAY-1999; 99US-0133459P.  
PR 02-JUN-1999; 99WO-US012252.  
PR 22-JUN-1999; 99US-0140650P.  
PR 22-JUN-1999; 99US-0140653P.  
PR 26-JUL-1999; 99US-0144758P.  
PR 26-JUL-1999; 99US-0145698P.  
PR 28-JUL-1999; 99US-0149395P.  
PR 31-AUG-1999; 99US-0151689P.  
PR 01-SEP-1999; 99WO-US020111.  
PR 15-SEP-1999; 99WO-US021090.  
PR 30-NOV-1999; 99WO-US028313.  
PR 01-DEC-1999; 99WO-US028301.  
PR 01-DEC-1999; 99WO-US028634.

PR 05-JAN-2000; 2000WO-US000219.  
XX (GETH ) GENENTECH INC.  
XX Ashkenazi AJ, Goddard A, Godowski PJ, Gurney AL, Hillan KJ;  
PI Marsters SA, Pan J, Pitti RM, Roy MA, Smith V, Stone DM;  
PI Watanabe CK, Wood WI;  
XX WPI; 2002-205567/26.  
DR N-PSDB; ABR40275.  
XX Thirty five nucleic acids encoding PRO polypeptides, useful for treating  
PT benign or malignant tumors, leukemias and lymphoid malignancies,  
PT inflammatory, angiogenic and immunologic disorders.  
XX  
XX Claim 61; Fig 44; 302pp; English.  
XX  
CC The present invention relates to the isolation of novel human PRO  
CC polypeptides and the polynucleotide sequences encoding them. The PRO  
CC polypeptides, agonists or antagonists or anti-PRO antibodies are useful for  
CC treating benign or malignant tumours (e.g. renal, kidney, bladder,  
CC breast, etc), leukemias and lymphoid malignancies, other disorders such  
CC as neuronal, glial, astrocytal, hypothalamic, glandular, macrophagal,  
CC stromal and blastocoelec disorders, inflammatory, immune and angiogenic  
CC disorders. The polynucleotide sequences are also useful in gene therapy.  
CC AAU86128-AAU86162 represent the human PRO polypeptides of the invention  
XX  
SQ Sequence 263 AA;  
  
Query Match 94.6%; Score 229; DB 5; Length 263;  
Best Local Similarity 100.0%; Pred. No. 1.5e-196;  
Matches 229; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 1 SGGCFWNGHLYREDQTSAPGLRCLNWLDAQSLASAPVSGAGNHSYCRNPDEPRGW 60  
Db 22 SGGCFWNGHLYREDQTSAPGLRCLNWLDAQSLASAPVSGAGNHSYCRNPDEPRGW 81  
  
QY 61 CYVSGEAGVPEKPCEDLRCPETTSQALPAFTTEIQEASGPGADVQVAFANALPARS 120  
Db 82 CYVSGEAGVPEKPCEDLRCPETTSQALPAFTTEIQEASGPGADVQVAFANALPARS 141  
  
QY 121 EAAAVQPVIGISQVRVNSKEKDLGLTGLVGLITMVMVIIAIGAGIILGYSYKRGKDLK 180  
Db 142 EAAAVQPVIGISQVRVNSKEKDLGLTGLVGLITMVMVIIAIGAGIILGYSYKRGKDLK 201  
  
QY 181 EQHDQKVCEREMQRTILPLSAFTNPTCEIVDEKTVVHTSQTPVDPQEG 229  
Db 202 EQHDQKVCEREMQRTILPLSAFTNPTCEIVDEKTVVHTSQTPVDPQEG 250  
  
RESULT 6  
AAB43237  
ID AAB43237 standard; protein; 263 AA.  
AC AAB43237;  
XX  
XX 08-FEB-2001 (first entry)  
XX  
XX Human ORF3001 polypeptide sequence SPQ ID NO:6002.  
DE  
XX  
XX Human; open reading frame; ORF; detection; cytostatic; hepatotropic;  
KW vulnery; antipsoiatric; antiparkinsonian; neurotic; neuroprotective;  
KW anticovulsant; osteopathic; antiarthritic; immunosuppressant; cardiant;  
KW immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;  
KW hypotensive; dermatological; immunosuppressive; antinflammatory;  
KW antiviral; antibacterial; antifungal; antirheumatic; antithyroid;  
KW antianemic; Gene therapy; cancer; proliferative disorder; hypertension;  
KW neurodegenerative disorder; osteoarthritis; graft vs host disease;  
KW cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;  
KW cholesterol ester storage; systemic lupus erythematosus; infection;  
KW severe combined immunodeficiency; malaria; autoimmune disorder; asthma;  
KW allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;  
KW bone damage; cartilage damage; antinflammatory disease; coagulation;

thrombosis; contraceptive.  
Homo sapiens.  
WO200058473-A2.  
05-OCT-2000.  
31-MAR-2000; 2000WO-US008621.  
31-MAR-1999; 99US-0127607P.  
02-APR-1999; 99US-0127636P.  
05-APR-1999; 99US-0127728P.  
30-MAR-2000; 2000US-00540763.  
(CURA-) CURAGEN CORP.  
Shimkets RA, Leach M;  
WPI; 2000-602362/57.  
N-PSDB; AAC77446.  
Novel nucleic acids and peptides derived from open reading frame X,  
useful for treating e.g. cancers, proliferative disorders,  
neurodegenerative disorders and cardiovascular disease.  
Claim 11; Page 5181-5182; 5507pp; English.  
AAC74446 to AAC7606 encode the proteins given in AAB40237 to AAB43397,  
which represent the human ORF open reading frames 1 to 3161. The ORF  
sequences have activities such as: cytostatic; hepatotropic; vulnary;  
antiparasitic; antiparkinsonian; nontropic; neuroprotective; osteopathic;  
anticonvulsant; antiarthritic; immunosuppressant; immunostimulant;  
cardiant; thrombolytic; coagulant; vasotropic; antidiabetic; hypotensive;  
dermatological; immunosuppressive; antineoplastic; antibacterial;  
antiviral; antifungal; antirheumatic; antithyroid; and antianaemic. The  
sequences can be used for determining the presence of or predisposition  
to, or preventing or treating pathological conditions associated with an  
ORF-associated disorder. The nucleic acids can be used to express ORF  
proteins in gene therapy vectors. The proteins and nucleic acids may be  
used to treat cancers, proliferative disorders, neurodegenerative  
disorders, osteoarthritis, graft vs host disease, cardiovascular disease,  
diabetes mellitus, hypertension, hypothyroidism, cholesterol ester  
storage, systemic lupus erythematosus, severe combined immunodeficiency  
(SCID), AIDS, viral, bacterial or fungal infection, malaria, autoimmune  
disorders, asthma, allergies, aplastic anaemia, burns, wounds, bone and  
cartilage damage, nocturnal haemoglobinuria, antineoplastic disease; to  
enhance coagulation; to inhibit thrombosis; and as a contraceptive  
Sequence 263 AA;  
Query Match 71.1%; Score 172; DB 3; Length 263;  
Best Local Similarity 100.0%; Pred. No. 1.5e-145; Indels 0; Gaps 0;  
Matches 172; Conservative 0; Mismatches 0;  
QY 58 GPWCYVSGEAGVPEKPCEDLRCPETTSQALPATTIIOEASGEGDADEVQVAPANALP 117  
Db 79 GPWCYVSGEAGVPEKPCEDLRCPETTSQALPATTIIOEASGEGDADEVQVAPANALP 138  
QY 118 ARSEAAVQPVIGISQVRMNSKEKDLGTIGYVLGIMVVIITAGIGILGYSYRGK 177  
Db 139 ARSEAAVQPVIGISQVRMNSKEKDLGTIGYVLGIMVVIITAGIGILGYSYRGK 198  
QY 178 DLKEQHDQKVCEREMQRTITLPLSAFTNPTCEIVDEKTVVHTSTQTPVDPQSG 229  
Db 199 DLKEQHDQKVCEREMQRTITLPLSAFTNPTCEIVDEKTVVHTSTQTPVDPQSG 250  
RESULT 7  
ABR40487  
ID ABR40487 standard; protein; 146 AA.  
XX  
AC ABR40487;

13-JUN-2003 (first entry)  
Human secreted protein #SEQ ID 237.  
Human; secreted protein; anti-HIV; nontropic; neuroprotective;  
antiangiogenic; immunosuppressive; immunomodulator; cytostatic; cardiant;  
hepatotropic; antiinflammatory; antiallergic; antidiabetic;  
gastrointestinal; antifertility; nephrotropic; virucide; hypotensive;  
vasotropic; dermatological; osteopathic; antiarthritic; antiparkinsonian;  
antisthmatic; antiparasitic; cerebroprotective; antibacterial;  
fungicide; antiparasitic; protein therapy; gene therapy; cancer;  
hyperproliferative disorder; leukaemia; autoimmune disorder;  
immunodeficiency; HIV; infertility; uterine disorders; endometriosis;  
amenorrhea; ocular disorder; neurological disorder; wound healing;  
Huntington's disease; gastrointestinal disorder; inflammatory disease;  
Crohn's disease; kidney disorder; renal failure; diabetes; obesity;  
cirrhosis; hepatitis; cardiovascular disorder; angina; osteoarthritis;  
rheumatoid arthritis; psoriasis; multiple sclerosis; asthma;  
respiratory disorder; infectious disease; chromosome identification;  
food additive; nutrition.  
Homo sapiens.  
WO200268628-A1.  
06-SEP-2002.  
21-FEB-2002; 2002WO-US005301.  
23-FEB-2001; 2001US-0270625P.  
12-JUL-2001; 2001US-0304417P.  
(HUMA-) HUMAN GENOME SCI INC.  
Rosen CA, Komatsoulis G, Baker KP, Piscella M, Moore PA, Wei P;  
Duan DR, Shi Y, Gupta R;  
WPI; 2002-750417/81.  
N-PSDB; ABZ82542.  
New human secreted proteins and nucleic acids, useful for preventing,  
treating, or ameliorating a medical condition, e.g. cancer, autoimmune  
disorder, HIV, Parkinson's disease, Alzheimer's disease, diabetes,  
obesity or cirrhosis.  
Claim 11; Page 800; 873pp; English.  
The invention relates to novel human secreted proteins and the genes  
encoding them. Genes and proteins of the invention may be useful for  
preventing, treating or ameliorating medical conditions e.g. by protein  
or gene therapy. These conditions include cancer and hyperproliferative  
disorders, immune cell proliferative disorders (e.g. leukaemia),  
autoimmune disorders, immunodeficiencies (e.g. HIV infection),  
infertility, placental and uterine disorders (e.g. endometriosis),  
amenorrhea, ocular disorders, neurological disorders (e.g. Huntington's  
disease), wound healing, gastrointestinal system disorders, particularly  
inflammatory diseases (e.g. Crohn's disease), kidney disorders (e.g.  
renal failure), diabetes, obesity, cirrhosis, hepatitis, cardiovascular  
disorders (e.g. angina), rheumatoid arthritis, osteoarthritis, psoriasis,  
multiple sclerosis, respiratory disorders (e.g. asthma), or infectious  
diseases caused by bacterial, parasitic, viral or fungal agents. The  
nucleic acids are also useful for chromosome identification, radiation  
hybrid mapping or long-range restriction mapping. The polypeptide,  
polynucleotide, agonist or antagonist may also be used as a food additive  
or preservative to increase or decrease storage capabilities, fat content  
or other nutritional components. The sequences given in records ABR40499-  
ABR40590 and ABZ82464-ABZ82611 represent human secreted proteins and the  
genes encoding them  
Sequence 146 AA;  
Query Match 45.5%; Score 110; DB 5; Length 146;

Best Local Similarity 100.0%; Pred. No. 2.9e-90; Mismatches 0; Indels 0; Gaps 0;  
Matches 110; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SGGCFWNGHLYREDQTSAPGLRCLNWLDAQSLASAPVSGAGNHSYCRNPDPRGPW 60  
DB 22 SGGCFWNGHLYREDQTSAPGLRCLNWLDAQSLASAPVSGAGNHSYCRNPDPRGPW 81

QY 61 CYSVGEAGVPEKPCEDLRCPETTSQALPAFTTIEQASEGPGADEVQVF 110  
DB 82 CYSVGEAGVPEKPCEDLRCPETTSQALPAFTTIEQASEGPGADEVQVF 131

RESULT 8  
ABR40561  
ID ABR40561 standard; protein; 146 AA.  
AC ABR40561;  
XX  
XX  
DT 13-JUN-2003 (first entry)  
DE Human secreted protein #SEQ ID 311.  
XX  
XX  
KW Human; secreted protein; anti-HIV; nontropic; neuroprotective;  
KW antiangiogenic; immunosuppressive; immunomodulator; cytostatic; cardiant;  
KW hepatotropic; antiinflammatory; antiallergic; antidiabetic;  
KW gastroenteric; antifertility; nephrotropic; virucide; hypotensive;  
KW vasotropic; dermatological; osteopathic; antiarthritic; antiparkinsonian;  
KW antiasthmatic; antispasmodic; cerebroprotective; antibacterial;  
KW fungicide; antiparasitic; protein therapy; gene therapy; cancer;  
KW hyperproliferative disorder; leukaemia; autoimmune disorder;  
KW immunodeficiency; HIV; infertility; uterine disorders; endometriosis;  
KW amenorrhea; ocular disorder; neurological disorder; wound healing;  
KW Huntington's disease; gastrointestinal disorder; inflammatory disease;  
KW Crohn's disease; kidney disorder; renal failure; diabetes; obesity;  
KW cirrhosis; hepatitis; cardiovascular disorder; angina; osteoarthritis;  
KW rheumatoid arthritis; psoriasis; multiple sclerosis; asthma;  
KW respiratory disorder; infectious disease; chromosome identification;  
KW food additive; nutrition.  
XX  
XX  
OS Homo sapiens.  
XX  
XX WO200268628-A1.  
XX  
XX  
PD 06-SEP-2002.  
XX  
XX 21-FEB-2002; 2002WO-US005301.  
XX  
XX 23-FEB-2001; 2001US-0270625P.  
XX  
XX 12-JUL-2001; 2001US-0304417P.  
XX  
XX (HUMA-) HUMAN GENOME SCI INC.  
XX  
XX  
XX Rosen CA, Komatsoulis G, Baker KP, Fiscella M, Moore PA, Wei P;  
XX Duan DR, Shi Y, Gupta R;  
XX WPI; 2002-750417/81.  
XX  
XX New human secreted proteins and nucleic acids, useful for preventing,  
XX treating, or ameliorating a medical condition, e.g. cancer, autoimmune  
XX disorder, HIV, Parkinson's disease, Alzheimer's disease, diabetes,  
XX obesity or cirrhosis.  
XX  
XX Claim 11; Page 850; 873pp; English.  
XX  
XX The invention relates to novel human secreted proteins and the genes  
XX encoding them. Genes and proteins of the invention may be useful for  
XX preventing, treating or ameliorating medical conditions e.g. by protein  
XX or gene therapy. These conditions include cancer and hyperproliferative  
XX disorders, immune cell proliferative disorders (e.g. leukaemia),  
XX autoimmune disorders, immunodeficiencies (e.g. HIV infection),  
XX infertility, placental and uterine disorders (e.g. endometriosis),  
XX amenorrhea, ocular disorders, neurological disorders (e.g. Huntington's  
XX disease), wound healing, gastrointestinal system disorders, particularly

CC inflammatory diseases (e.g. Crohn's disease), kidney disorders (e.g.  
CC renal failure), diabetes, obesity, cirrhosis, hepatitis, cardiovascular  
CC disorders (e.g. angina), rheumatoid arthritis, osteoarthritis, psoriasis,  
CC multiple sclerosis, respiratory disorders (e.g. asthma), or infectious  
CC diseases caused by bacterial, parasitic, viral or fungal agents. The  
CC nucleic acids are also useful for chromosome identification, radiation  
CC hybrid mapping or long-range restriction mapping. The polypeptide,  
CC polynucleotide, agonist or antagonist may also be used as a food additive  
CC or preservative to increase or decrease storage capabilities, fat content  
CC or other nutritional components. The sequences given in records ABR40409-  
CC ABR40590 and ABR282464-ABR282461 represent human secreted proteins and the  
CC genes encoding them  
XX  
SQ Sequence 146 AA;  
Query Match 45.5%; Score 110; DB 5; Length 146;  
Best Local Similarity 100.0%; Pred. No. 2.9e-90; Mismatches 0; Indels 0; Gaps 0;  
Matches 110; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SGGCFWNGHLYREDQTSAPGLRCLNWLDAQSLASAPVSGAGNHSYCRNPDPRGPW 60  
DB 22 SGGCFWNGHLYREDQTSAPGLRCLNWLDAQSLASAPVSGAGNHSYCRNPDPRGPW 81

QY 61 CYSVGEAGVPEKPCEDLRCPETTSQALPAFTTIEQASEGPGADEVQVF 110  
DB 82 CYSVGEAGVPEKPCEDLRCPETTSQALPAFTTIEQASEGPGADEVQVF 131

RESULT 9  
AAM93748  
ID AAM93748 standard; protein; 263 AA.  
XX  
XX AAM93748;  
AC AAM93748;  
XX  
XX 06-NOV-2001 (first entry)  
DT  
XX  
XX Human polypeptide, SEQ ID NO: 3727.  
DE  
XX  
XX Human; full length cDNA; cDNA synthesis; oligo-capping.  
XX  
XX Homo sapiens.  
XX  
XX EP1130094-A2.  
XX  
XX 05-SEP-2001.  
XX  
XX 07-JUL-2000; 2000EP-00114089.  
XX  
XX 08-JUL-1999; 99JP-00194486.  
XX  
XX 11-JAN-2000; 2000JP-00118774.  
XX  
XX 02-MAY-2000; 2000JP-00183765.  
XX  
XX (HELI-) HELIX RES INST.  
XX  
XX Ota T, Nishikawa T, Isogai T, Hayashi K, Ishii S, Kawai Y;  
XX Wakamatsu A, Sugiyama T, Nagai K, Kojima S, Otsuki T, Koga H;  
XX WPI; 2001-524255/58.  
XX  
XX N-PSDB; AAK94700.  
XX  
XX 830 Primers useful for synthesizing full length cDNA clones and their use  
XX in genetic manipulation.  
XX  
XX Claim 8; SEQ ID NO 3727; 1380pp + Sequence Listing; English.  
XX  
XX The invention relates to primers for synthesizing full length cDNA  
XX clones. 830 cDNA molecules encoding a human protein have been isolated  
XX and nucleotide sequences of 5' and 3' ends of the cDNA molecules have  
XX been determined. Primers for synthesizing the full length cDNA are useful  
XX for clarifying the function of the protein encoded by the cDNA. The full  
XX length clones were obtained by construction of full length enriched cDNA  
XX libraries that were synthesised by the oligo-capping method. The primers  
XX enable the production of the full length cDNA easily without any special



CC methods. The present sequence is a polypeptide encoded by a full length  
CC human cDNA of the invention. Note: The sequence data for this patent did  
CC not form part of the printed specification, but was obtained in CD-ROM  
CC format directly from EPO

XX SQ Sequence 263 AA;  
Query Match 41.7%; Score 101; DB 4; Length 263;  
Best Local Similarity 100.0%; Pred. No. 5.4e-82;  
Matches 101; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 142 KDLGLGVVLTGTTMMVHIIAGIILGYSYKRGKDLKEQHDQKVCEREMQRTPLPSA 201  
DB 163 KDLGLGVVLTGTTMMVHIIAGIILGYSYKRGKDLKEQHDQKVCEREMQRTPLPSA 222  
QY 202 FNTPTCEIVDEKTVVHTSQTPTVDPQEGSTPLMGQAGTPGA 242  
DB 223 FNTPTCEIVDEKTVVHTSQTPTVDPQEGSTPLMGQAGTPGA 263

RESULT 10  
AAM18800  
ID AAM18800 standard; protein; 66 AA.  
XX AC AAM18800;  
XX DT 12-OCT-2001 (first entry)  
XX DE Peptide #5234 encoded by probe for measuring cervical gene expression.  
XX KW Probe; human; microarray; gene expression; cervical epithelial cell;  
XX KW cervical cancer.  
XX OS Homo sapiens.  
XX XX WC200157278-A2.  
XX XX 09-AUG-2001.  
XX XX 30-JAN-2001; 2001WO-US000670.  
XX XX 04-FEB-2000; 2000US-0180312P.  
XX XX 26-MAY-2000; 2000US-0207456P.  
XX XX 30-JUN-2000; 2000US-00608408.  
XX XX 03-AUG-2000; 2000US-00632366.  
XX XX 21-SEP-2000; 2000US-0234687P.  
XX XX 27-SEP-2000; 2000US-0236359P.  
XX XX 04-OCT-2000; 2000GB-00024263.  
XX XX (MOLE-) MOLECULAR DYNAMICS INC.  
XX XX Penn SG, Hanzel DK, Chen W, Rank DR;  
XX WPI; 2001-488901/53.

XX Human genome-derived single exon nucleic acid probes useful for analyzing  
XX PT gene expression in human cervical epithelial cells.  
XX XX Claim 27; SEQ ID NO 23626; 487pp; English.

XX The present invention relates to human single exon nucleic acid probes  
XX (SENP; see AAI10068-AAT28459). The present sequence is a peptide encoded  
XX by one such probe. The SENPs are derived from human HeLa cells. The SENPs  
XX measuring human gene expression in a sample derived from human cervical  
XX epithelial cells. By measuring gene expression, the probes are therefore  
XX useful in grading and/or staging of diseases of the cervix, notably  
XX cervical cancer. Note: The sequence data for this patent did not form  
XX part of the printed specification, but was obtained in electronic format  
XX directly from WIPO at ftp.wipo.int/pub/published\_pct\_sequences  
XX SQ Sequence 66 AA;

Query Match 27.3%; Score 66; DB 4; Length 66;  
Best Local Similarity 100.0%; Pred. No. 3.5e-51;  
Matches 66; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 83 TTSQLPAFTTEIQEASGPGADEVQVFAPANALPARSEAAVQPVIGISQVRVNSKEK 142  
DB 1 TTSQLPAFTTEIQEASGPGADEVQVFAPANALPARSEAAVQPVIGISQVRVNSKEK 60  
QY 143 KDLGTL 148  
|||

Query Match 27.3%; Score 66; DB 4; Length 66;  
Best Local Similarity 100.0%; Pred. No. 3.5e-51;  
Matches 66; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 83 TTSQLPAFTTEIQEASGPGADEVQVFAPANALPARSEAAVQPVIGISQVRVNSKEK 142  
DB 1 TTSQLPAFTTEIQEASGPGADEVQVFAPANALPARSEAAVQPVIGISQVRVNSKEK 60  
QY 143 KDLGTL 148  
DB 61 KDLGTL 66  
|||

RESULT 11  
ABB37905  
ID ABB37905 standard; peptide; 66 AA.  
XX AC ABB37905;  
XX DT 04-FEB-2002 (first entry)  
XX DE Peptide #5411 encoded by human foetal liver single exon probe.  
XX KW Human; foetal liver; gene expression; single exon nucleic acid probe.  
XX OS Homo sapiens.  
XX XX WO200157277-A2.  
XX XX 09-AUG-2001.  
XX XX 30-JAN-2001; 2001WO-US000669.  
XX XX 04-FEB-2000; 2000US-0180312P.  
XX XX 26-MAY-2000; 2000US-0207456P.  
XX XX 30-JUN-2000; 2000US-00608408.  
XX XX 03-AUG-2000; 2000US-00632366.  
XX XX 21-SEP-2000; 2000US-0234687P.  
XX XX 27-SEP-2000; 2000US-0236359P.  
XX XX 04-OCT-2000; 2000GB-00024263.  
XX XX (MOLE-) MOLECULAR DYNAMICS INC.  
XX XX Penn SG, Hanzel DK, Chen W, Rank DR;  
XX WPI; 2001-483447/52.

XX Human genome-derived single exon nucleic acid probes useful for analyzing  
XX PT gene expression in human foetal liver.  
XX XX Claim 27; SEQ ID NO 30540; 639pp + Sequence Listing; English.  
XX The invention relates to a single exon nucleic acid probe for measuring  
XX human gene expression in a sample derived from human foetal liver. The  
XX single exon nucleic acid probes may be used for predicting, measuring and  
XX displaying gene expression in samples derived from human foetal liver. The  
XX present sequence is a peptide encoded by a single exon nucleic acid probe  
XX of the invention. Note: The sequence data for this patent did not form  
XX part of the printed specification, but was obtained in electronic format  
XX directly from WIPO at ftp.wipo.int/pub/published\_pct\_sequences  
XX SQ Sequence 66 AA;

Query Match 27.3%; Score 66; DB 4; Length 66;  
Best Local Similarity 100.0%; Pred. No. 3.5e-51;  
Matches 66; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 83 TTSQLPAFTTEIQEASGPGADEVQVFAPANALPARSEAAVQPVIGISQVRVNSKEK 142  
DB 1 TTSQLPAFTTEIQEASGPGADEVQVFAPANALPARSEAAVQPVIGISQVRVNSKEK 60  
QY 143 KDLGTL 148  
|||

Db 61 KDLGTL 66

RESULT 12  
AAM31314  
ID. AAM31314 standard; protein; 66 AA.  
XX  
AC AAM31314;  
XX  
DT 17-OCT-2001 (first entry)  
XX  
DE Peptide #5351 encoded by probe for measuring placental gene expression.  
XX  
KW Probe; microarray; human; placenta; antenatal diagnosis;  
KW genetic disorder.  
XX  
OS Homo sapiens.  
XX  
FN WO200157272-A2.  
XX  
PD 09-AUG-2001.  
XX  
PF 30-JAN-2001; 2001WO-US000663.  
XX  
PR 04-FEB-2000; 2000US-0180312P.  
PR 26-MAY-2000; 2000US-0207456P.  
PR 30-JUN-2000; 2000US-00608408.  
PR 03-AUG-2000; 2000US-00632366.  
PR 21-SEP-2000; 2000US-0234687P.  
PR 27-SEP-2000; 2000US-0236359P.  
PR 04-OCT-2000; 2000GB-00024263.  
XX  
PA (MOLE-) MOLECULAR DYNAMICS INC.  
XX  
PI Penn SG, Hanzel DK, Chen W, Rank DR;  
XX  
XX WPI; 2001-488897/53.  
XX  
PT Human genome-derived single exon nucleic acid probes useful for analyzing  
PT gene expression in human placenta.  
XX  
PS Claim 27; SEQ ID NO 31583; 654pp; English.  
XX  
CC The present invention relates to single exon nucleic acid probes (SENP:  
CC see AA31315-AA157546). The present sequence is a peptide encoded by one  
CC such probe. The probes are useful for producing a microarray for  
CC predicting, measuring and displaying gene expression in samples derived  
CC from human placenta. The probes are useful for antenatal diagnosis of  
CC human genetic disorders  
XX  
SQ Sequence 66 AA;  
Query Match 27.3%; Score 66; DB 4; Length 66;  
Best Local Similarity 100.0%; Pred. No. 3.5e-51;  
Matches 66; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 83 TTSQLPAPFTTIOEASEGFGADEVQVFAPANALPARSEAAAVQPVIGISQVRMNSKEK 142  
Db 1 TTSQLPAPFTTIOEASEGFGADEVQVFAPANALPARSEAAAVQPVIGISQVRMNSKEK 60  
QY 143 KDLGTL 148  
Db 61 KDLGTL 66

RESULT 13  
ABB23159  
ID ABB23159 standard; protein; 66 AA.  
XX  
AC ABB23159;  
XX  
DT 23-JAN-2002 (first entry)  
XX  
DE Human bone marrow expressed probe encoded protein SEQ ID NO: 31343.  
XX Human; bone marrow expressed exon; gene expression analysis; probe;  
KW microarray; cancer; leukaemia; lymphoma; myeloma.

DE XX  
XX Protein #5158 encoded by probe for measuring heart cell gene expression.  
KW Human; gene expression; heart; microarray; vascular system;  
KW cardiovascular disease; hypertension; cardiac arrhythmia;  
KW congenital heart disease.  
XX  
OS Homo sapiens.  
XX  
PN WO200157274-A2.  
XX  
PD 09-AUG-2001.  
XX  
PF 30-JAN-2001; 2001WO-US000666.  
XX  
PR 04-FEB-2000; 2000US-0180312P.  
PR 26-MAY-2000; 2000US-0207456P.  
PR 30-JUN-2000; 2000US-00608408.  
PR 03-AUG-2000; 2000US-00632366.  
PR 21-SEP-2000; 2000US-0234687P.  
PR 27-SEP-2000; 2000US-0236359P.  
PR 04-OCT-2000; 2000GB-00024263.  
XX  
PA (MOLE-) MOLECULAR DYNAMICS INC.  
XX  
PI Penn SG, Hanzel DK, Chen W, Rank DR;  
XX  
XX WPI; 2001-488899/53.  
XX  
PT Single exon nucleic acid probes for analyzing gene expression in human  
PT hearts.  
XX  
PS Claim 15; SEQ ID NO 24929; 530pp; English.  
XX  
CC The present invention relates to single exon nucleic acid probes for  
CC measuring human gene expression in a sample derived from human heart (see  
CC ABA21535-ABA41305). The present sequence is a protein encoded by one such  
CC probe. The probes may be used for predicting, measuring and displaying  
CC gene expression in samples derived from the human heart via microarrays.  
CC By measuring gene expression, the probes are useful for predicting,  
CC diagnosing, grading, staging, monitoring and prognosing diseases of the  
CC human heart and vascular system e.g. cardiovascular disease,  
CC hypertension, cardiac arrhythmias and congenital heart disease. Note: The  
CC sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences  
XX  
SQ Sequence 66 AA;  
Query Match 27.3%; Score 66; DB 4; Length 66;  
Best Local Similarity 100.0%; Pred. No. 3.5e-51;  
Matches 66; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 83 TTSQLPAPFTTIOEASEGFGADEVQVFAPANALPARSEAAAVQPVIGISQVRMNSKEK 142  
Db 1 TTSQLPAPFTTIOEASEGFGADEVQVFAPANALPARSEAAAVQPVIGISQVRMNSKEK 60  
QY 143 KDLGTL 148  
Db 61 KDLGTL 66

RESULT 14  
AAM71037  
ID AAM71037 standard; protein; 66 AA.  
XX  
AC AAM71037;  
XX  
DT 06-NOV-2001 (first entry)  
XX  
DE Human bone marrow expressed probe encoded protein SEQ ID NO: 31343.  
XX Human; bone marrow expressed exon; gene expression analysis; probe;  
KW microarray; cancer; leukaemia; lymphoma; myeloma.

PR 30-JUN-2000; 2000US-00608408.  
 PR 03-AUG-2000; 2000US-00832366.  
 PR 21-SEP-2000; 2000US-0234687P.  
 PR 27-SEP-2000; 2000US-0236359P.  
 PR 04-OCT-2000; 2000GB-00024263.  
 XX (MOLE-) MOLECULAR DYNAMICS INC.  
 XX PA Penn SG, Hanzel DK, Chen W, Rank DR;  
 XX PI MPI; 2001-483446/52.  
 XX DR Single exon nucleic acid probes for analyzing gene expression in human  
 XX PT brains.  
 XX PS Example 4; SEQ ID NO 30642; 650pp + Sequence Listing; English.  
 XX CC The present invention provides a number of single exon nucleic acid  
 XX CC probes which are derived from genomic sequences expressed in the human  
 XX CC brain. They can be used to measure gene expression in brain cell samples,  
 XX CC which may enable the diagnosis and improved treatment of nervous system  
 XX CC diseases such as Alzheimer's disease, multiple sclerosis, schizophrenia,  
 XX CC epilepsy and cancers. The present sequence is a protein encoded by one of  
 XX CC the probes of the invention  
 XX SQ Sequence 66 AA;  
 Query Match 27.3%; Score 66; DB 4; Length 66;  
 Best Local Similarity 100.0%; Pred. No. 3.5e-51;  
 Matches 66; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 83 TTSCALPAFTTETQEASEGGADEVQVFAPANALPARSEAAAVQPVIGISQVRMNSKEK 142  
 DB 1 TTSCALPAFTTETQEASEGGADEVQVFAPANALPARSEAAAVQPVIGISQVRMNSKEK 60  
 QY 143 KDLGTL 148  
 DB 61 KDLGTL 66  
 RESULT 16  
 ABG52752  
 ID ABG52752 standard; peptide; 66 AA.  
 XX AC ABG52752;  
 XX DT 25-FEB-2003 (first entry)  
 XX DE Human liver peptide, SEQ ID No 31400.  
 XX KW Human; liver; cirrhosis; hyperlipoproteinaemia; hyperlipidaemia;  
 XX KW hypercholesterolaemia; coronary heart disease.  
 XX OS Homo sapiens.  
 XX PN WO200157273-A2.  
 XX PD 09-AUG-2001.  
 XX PF 30-JAN-2001; 2001WO-US000664.  
 XX \* 04-FEB-2000; 2000US-0180312P.  
 XX PR 26-MAY-2000; 2000US-0207456P.  
 XX PR 30-JUN-2000; 2000US-00608408.  
 XX PR 03-AUG-2000; 2000US-00632366.  
 XX PR 21-SEP-2000; 2000US-0234687P.  
 XX PR 27-SEP-2000; 2000US-0236359P.  
 XX PR 04-OCT-2000; 2000GB-00024263.  
 XX (MOLE-) MOLECULAR DYNAMICS INC.  
 XX PA Penn SG, Hanzel DK, Chen W, Rank DR;  
 XX PI MPI; 2001-483446/52.  
 XX DR Single exon nucleic acid probes for analyzing gene expression in human  
 XX PT brains.  
 XX PS Example 4; SEQ ID NO 30642; 650pp + Sequence Listing; English.  
 XX CC The present invention provides a number of single exon nucleic acid  
 XX CC probes which are derived from genomic sequences expressed in the human  
 XX CC brain. They can be used to measure gene expression in brain cell samples,  
 XX CC which may enable the diagnosis and improved treatment of nervous system  
 XX CC diseases such as Alzheimer's disease, multiple sclerosis, schizophrenia,  
 XX CC epilepsy and cancers. The present sequence is a protein encoded by one of  
 XX CC the probes of the invention  
 XX SQ Sequence 66 AA;  
 Query Match 27.3%; Score 66; DB 4; Length 66;  
 Best Local Similarity 100.0%; Pred. No. 3.5e-51;  
 Matches 66; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 83 TTSCALPAFTTETQEASEGGADEVQVFAPANALPARSEAAAVQPVIGISQVRMNSKEK 142  
 DB 1 TTSCALPAFTTETQEASEGGADEVQVFAPANALPARSEAAAVQPVIGISQVRMNSKEK 60  
 QY 143 KDLGTL 148  
 DB 61 KDLGTL 66  
 RESULT 15  
 AAM58537  
 ID AAM58537 standard; protein; 66 AA.  
 XX AC AAM58537;  
 XX DT 05-NOV-2001 (first entry)  
 XX DE Human brain expressed single exon probe encoded protein SEQ ID NO: 30642.  
 XX KW Human; brain expressed exon; gene expression analysis; probe; microarray;  
 XX KW Alzheimer's disease; multiple sclerosis; schizophrenia; epilepsy; cancer.  
 XX OS Homo sapiens.  
 XX PN WO200157275-A2.  
 XX PD 09-AUG-2001.  
 XX PF 30-JAN-2001; 2001WO-US000667.  
 XX \* 04-FEB-2000; 2000US-0180312P.  
 XX PR 26-MAY-2000; 2000US-0207456P.

XX Homo sapiens.  
 XX WO200157276-A2.  
 XX 09-AUG-2001.  
 XX 30-JAN-2001; 2001WO-US000668.  
 XX 04-FEB-2000; 2000US-0180312P.  
 XX 26-MAY-2000; 2000US-0207456P.  
 XX 30-JUN-2000; 2000US-00608408.  
 XX 03-AUG-2000; 2000US-00632366.  
 XX 21-SEP-2000; 2000US-0234687P.  
 XX 27-SEP-2000; 2000US-0236359P.  
 XX 04-OCT-2000; 2000GB-00024263.  
 XX (MOLE-) MOLECULAR DYNAMICS INC.  
 XX PA Penn SG, Hanzel DK, Chen W, Rank DR;  
 XX PI MPI; 2001-488900/53.  
 XX DR Human genome-derived single exon nucleic acid probes useful for analyzing  
 XX PT gene expression in human bone marrow.  
 XX PS Example 4; SEQ ID NO 31343; 658pp + Sequence Listing; English.  
 XX CC The present invention provides a number of single exon nucleic acid  
 XX CC probes which are derived from genomic sequences expressed in the human  
 XX CC bone marrow. They can be used to measure gene expression in bone marrow  
 XX CC samples, which may enable the improved diagnosis and treatment of cancers  
 XX CC such as lymphoma, leukaemia and myeloma. The present sequence is a  
 XX CC protein encoded by one of the probes of the invention  
 XX SQ Sequence 66 AA;  
 Query Match 27.3%; Score 66; DB 4; Length 66;  
 Best Local Similarity 100.0%; Pred. No. 3.5e-51;  
 Matches 66; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 83 TTSCALPAFTTETQEASEGGADEVQVFAPANALPARSEAAAVQPVIGISQVRMNSKEK 142  
 DB 1 TTSCALPAFTTETQEASEGGADEVQVFAPANALPARSEAAAVQPVIGISQVRMNSKEK 60  
 QY 143 KDLGTL 148  
 DB 61 KDLGTL 66  
 RESULT 15  
 AAM58537  
 ID AAM58537 standard; protein; 66 AA.  
 XX AC AAM58537;  
 XX DT 05-NOV-2001 (first entry)  
 XX DE Human brain expressed single exon probe encoded protein SEQ ID NO: 30642.  
 XX KW Human; brain expressed exon; gene expression analysis; probe; microarray;  
 XX KW Alzheimer's disease; multiple sclerosis; schizophrenia; epilepsy; cancer.  
 XX OS Homo sapiens.  
 XX PN WO200157275-A2.  
 XX PD 09-AUG-2001.  
 XX PF 30-JAN-2001; 2001WO-US000667.  
 XX \* 04-FEB-2000; 2000US-0180312P.  
 XX PR 26-MAY-2000; 2000US-0207456P.

DR WPI; 2001-488998/53.  
XX Human genome-derived single exon nucleic acid probes useful for analyzing  
PT gene expression in human adult liver.  
XX  
XX  
PS Claim 27; SEQ ID NO 31400; 658pp; English.  
XX  
XX The invention relates to a single exon nucleic acid probe (SENP) (I) for  
CC measuring human gene expression in a sample derived from human adult  
CC liver, comprising one of 13109 defined nucleotide sequences given in the  
CC specification (or complements/ fragments). The probe hybridises at high  
CC stringency to a nucleic acid molecule expressed in the human adult liver.  
CC (I) may be used for predicting, measuring and displaying gene expression  
CC in samples derived from human adult liver. The genes identified may be  
CC involved in genetic liver diseases such as cirrhosis,  
CC hyperlipoproteinaemia, hyperlipidaemia and hypercholesterolaemia which is  
CC associated with coronary heart disease. ABG47348-ABG59930 represent human  
CC liver single exon encoded peptides of the invention. Note: The sequence  
CC information for this patent does not appear in the printed specification  
CC but was obtained in electronic format directly from WIPO at  
CC ftp.wipo.int/pub/published\_pct\_sequences  
XX  
SQ Sequence 66 AA;  
Query Match 27.3%; Score 66; DB 4; Length 66;  
Best Local Similarity 100.0%; Pred. No. 3.5e-51;  
Matches 66; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 83 TTSQLPAFTTETIQEASGPGADEVQVFAPANALPARSEAAAVQVIGISQVRVNSKEK 142  
DB 1 TTSQLPAFTTETIQEASGPGADEVQVFAPANALPARSEAAAVQVIGISQVRVNSKEK 60  
QY 143 KDLGTL 148  
DB 61 KDLGTL 66  
RESULT 17  
ABG40828  
ID ABG40828 standard; peptide; 66 AA.  
XX  
XX ABG40828;  
AC  
XX  
DT 19-AUG-2002 (first entry)  
XX  
DE Human peptide encoded by genome-derived single exon probe SEQ ID 30493.  
XX  
XX Human; single exon probe; asthma; lung cancer; COPD; ILD;  
KW chronic obstructive pulmonary disease; interstitial lung disease;  
KW familial idiopathic pulmonary fibrosis; neurofibromatosis;  
KW tuberous sclerosis; Gaucher's disease; Niemann-Pick disease;  
KW Hermansky-Pudlak syndrome; sarcoidosis; pulmonary haemosiderosis;  
KW pulmonary histiocytosis; lymphangioleiomyomatosis; Karageners syndrome;  
KW pulmonary alveolar proteinosis; fibrocystic pulmonary dysplasia;  
KW primary ciliary dyskinesia; pulmonary hypertension;  
KW hyaline membrane disease.  
XX  
OS Homo sapiens.  
XX  
FN WO200186003-A2.  
XX  
PD 15-NOV-2001.  
XX  
PF 30-JAN-2001; 2001WO-US000665.  
XX  
PR 04-FEB-2000; 2000US-0180312P.  
PR 26-MAY-2000; 2000US-0207456P.  
PR 30-JUN-2000; 2000US-00608408.  
PR 03-AUG-2000; 2000US-00632366.  
PR 21-SEP-2000; 2000US-0234687P.  
PR 27-SEP-2000; 2000US-0236359P.  
PR 04-OCT-2000; 2000GB-00024263.  
XX  
PA (MOLE-) MOLECULAR DYNAMICS INC.  
XX Penn SG, Hanzel DK, Chen W, Rank DR;  
XX WPI; 2002-114183/15.  
XX  
XX Spatially-addressable set of single exon nucleic acid probes, used to  
PT measure gene expression in human lung samples.  
XX  
XX Claim 27; SEQ ID NO 30493; 634pp; English.  
XX  
XX The invention relates to a spatially-addressable set of single exon  
CC nucleic acid probes for measuring gene expression in a sample derived  
CC from human lung comprising single exon nucleic acid probes having one of  
CC 12614 nucleic acid sequences mentioned in the specification, or their  
CC complements or the 12387 open reading frames derived from the 12614  
CC probes. Also included are a microarray comprising the novel set of probes  
CC ; the novel set of probes which hybridise at high stringency to a nucleic  
CC acid expressed in the human lung; measuring gene expression in a sample  
CC derived from human lung, comprising (a) contacting the array with a  
CC collection of detectably labeled nucleic acids derived from human lung  
CC mRNA, and (b) measuring the label detectably bound to each probe of the  
CC array; identifying exons in a eukaryotic genome, comprising (a)  
CC algorithmically predicting at least one exon from genomic sequences of  
CC the eukaryote; and (b) detecting specific hybridisation of detectably  
CC labeled nucleic acids from eukaryote lung mRNA, to a single exon probe,  
CC having a fragment identical to the predicted exon, the probe is included  
CC in the above mentioned microarray; assigning exons to a single gene,  
CC comprising (a) identifying exons from genomic sequence by the method  
CC above and (b) measuring the expression of each of the exons in several  
CC tissues and/or cell types using hybridisation to a single exon  
CC microarrays having a probe with the exon, where a common pattern of  
CC expression of the exons in the tissues and/or cell types indicates that  
CC the exons should be assigned to a single gene; a peptide comprising one  
CC of 12011 sequences, mentioned in the specification, or encoded by the  
CC probes/open reading frames (ORF). The probes are used for gene expression  
CC analysis, and for identifying exons in a gene, particularly using human  
CC lung derived mRNA and for the study of lung diseases such as asthma, lung  
CC cancer, chronic obstructive pulmonary disease (COPD), interstitial lung  
CC disease (ILD), familial idiopathic pulmonary fibrosis, neurofibromatosis,  
CC tuberous sclerosis, Gaucher's disease, Niemann-Pick disease, Hermansky-  
CC Pudlak syndrome, sarcoidosis, pulmonary haemosiderosis, pulmonary  
CC histiocytosis, lymphangioleiomyomatosis, pulmonary alveolar proteinosis,  
CC Karageners syndrome, fibrocystic pulmonary dysplasia, primary ciliary  
CC dyskinesia, pulmonary hypertension and hyaline membrane disease. The  
CC present sequence is a peptide/protein encoded by a single exon probe of  
CC the invention. Note: The sequence data for this patent did not form part  
CC of the printed specification, but was obtained in electronic format  
CC directly from WIPO at ftp.wipo.int/pub/published\_pct\_sequences  
XX  
SQ Sequence 66 AA;  
Query Match 27.3%; Score 66; DB 5; Length 66;  
Best Local Similarity 100.0%; Pred. No. 3.5e-51;  
Matches 66; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 83 TTSQLPAFTTETIQEASGPGADEVQVFAPANALPARSEAAAVQVIGISQVRVNSKEK 142  
DB 1 TTSQLPAFTTETIQEASGPGADEVQVFAPANALPARSEAAAVQVIGISQVRVNSKEK 60  
QY 143 KDLGTL 148  
DB 61 KDLGTL 66  
RESULT 18  
AA105220  
ID AA105220 standard; protein; 286 AA.  
XX  
XX AA105220;  
AC  
XX  
DT 17-JUN-1999 (first entry)  
XX

DE Kringle1 protein sequence.  
 XX Kringle1; cancer; inflammation; autoimmunity; rheumatoid arthritis; AIDS;  
 KW CNS inflammation; cerebellar degeneration; Alzheimer's disease; asthma;  
 KW Parkinson's disease; multiple sclerosis; amyotrophic lateral sclerosis;  
 KW neurological abnormality; ischaemia reperfusion injury; ischaemic injury;  
 KW cardiovascular disease; kidney disease; liver disease; aplastic anaemia;  
 KW myocardial infarction; hypotension; hypertension; allergy; infection;  
 KW myelodysplastic syndrome; haematologic abnormality; diagnosis; therapy;  
 XX male pattern baldness.  
 XX Homo sapiens.  
 OS  
 PN WO9906548-A2.  
 XX  
 PD 11-FEB-1999.  
 XX  
 PF 31-JUL-1998; 98WO-IB001222.  
 XX  
 PR 01-AUG-1997; 97US-00905135.  
 XX  
 PS (GEST ) GENSET.  
 XX  
 PI Dumas Milne Edwards J, Duclert A, Lacroix B;  
 XX  
 DR WPI; 1999-153778/13.  
 DR N-PSDB; AAX41230.  
 XX  
 PT New nucleic acids encoding human secreted proteins - obtained from cDNA  
 PT libraries prepared from e.g. liver, ovary, brain, prostate, kidney, lung,  
 PT umbilical cord, placenta and colon tissue.  
 XX  
 PS Claim 27; Page 744; 824pp; English.  
 XX  
 CC AAX41094 to AAX41347 represent 5' expressed sequence tags (ESTs) for  
 CC human secreted proteins, and encode the proteins given in AAY1261 to  
 CC AAY12614 respectively. The proteins given represent the signal peptide  
 CC and an N-terminal fragment of a secreted protein. The nucleic acid  
 CC sequences can be used for producing secreted human gene products. They  
 CC can also be used to develop products for diagnosis and therapy. The  
 CC proteins obtained may have cytokine activity, cell  
 CC proliferation/differentiation activity, haematopoiesis regulating  
 CC activity, tissue growth regulating activity, reproductive hormone  
 CC regulating activity, chemotactic/chemokinetic activity, anti-inflammatory  
 CC thrombolytic activity, receptor/ligand activity, anti-infectory  
 CC activity, tumour inhibition activity or other activities. The products  
 CC can be used in forensic, gene therapy and chromosome mapping procedures.  
 CC The sequences can also be used for obtaining corresponding promoter  
 CC sequences. The nucleic acids encoding the signal peptide can be used for  
 CC directing extracellular secretion of a polypeptide or the insertion of a  
 CC polypeptide into a membrane, or importing a polypeptide into a cell  
 XX  
 SQ Sequence 55 AA;  
 Query Match 14.0%; Score 34; DB 2; Length 55;  
 Best Local Similarity 100.0%; Pred. No. 1.3e-22;  
 Matches 34; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY \* 1 SGGCFWNGHLYREDQTSPPAPGLRCLNWLDAQSG 34  
 DB 22 SGGCFWNGHLYREDQTSPPAPGLRCLNWLDAQSG 55  
 RESULT 20  
 AAY12615  
 ID AAY12615 standard; protein; 56 AA.  
 XX  
 AC AAY12615;  
 XX  
 DT 22-JUN-1999 (first entry)  
 XX  
 DE Human 5' EST secreted protein SEQ ID NO: 280 from WO 9906553.  
 XX  
 KW Human; secreted protein; EST; expressed sequence tag; diagnosis;  
 KW forensic; gene therapy; chromosome mapping; signal peptide;

DE Kringle1 protein sequence.  
 XX Kringle1; cancer; inflammation; autoimmunity; rheumatoid arthritis; AIDS;  
 KW CNS inflammation; cerebellar degeneration; Alzheimer's disease; asthma;  
 KW Parkinson's disease; multiple sclerosis; amyotrophic lateral sclerosis;  
 KW neurological abnormality; ischaemia reperfusion injury; ischaemic injury;  
 KW cardiovascular disease; kidney disease; liver disease; aplastic anaemia;  
 KW myocardial infarction; hypotension; hypertension; allergy; infection;  
 KW myelodysplastic syndrome; haematologic abnormality; diagnosis; therapy;  
 XX male pattern baldness.  
 XX Homo sapiens.  
 OS  
 PN WO9911788-A1.  
 XX  
 PD 11-MAR-1999.  
 XX  
 PF 02-SEP-1998; 98WO-US018270.  
 XX  
 PR 02-SEP-1997; 97US-0056032P.  
 PR 01-SEP-1998; 98US-00144889.  
 XX  
 PS (SMIK ) SMITHKLINE BEECHAM CORP.  
 XX  
 PI Albane EF, Kixly KK;  
 XX  
 DR WPI; 1999-214707/18.  
 DR N-PSDB; AAX28355.  
 XX  
 PT New kringle1 polypeptides and polynucleotides.  
 XX  
 PS Claim 14; Page 33; 42pp; English.  
 XX  
 CC This sequence is a Kringle1 polypeptide of the invention. The kringle1  
 CC polypeptides (I) are used to screen for agonists and antagonists.  
 CC Agonists are used to treat subjects in need of enhanced activity or  
 CC expression of (I). Antagonists are used to treat subjects having need to  
 CC inhibit the activity or expression of (I). The methods can be used to  
 CC treat conditions such as cancer, inflammation, autoimmunity, allergy,  
 CC asthma, rheumatoid arthritis, CNS inflammation, cerebellar degeneration,  
 CC Alzheimer's disease, Parkinson's disease, multiple sclerosis,  
 CC amyotrophic lateral sclerosis, head injury damage and other neurological  
 CC abnormalities, ischaemia reperfusion injury, cardiovascular disease,  
 CC kidney disease, liver disease, ischaemic injury, myocardial infarction,  
 CC hypotension, hypertension, AIDS, myelodysplastic syndromes and other  
 CC haematologic abnormalities, aplastic anaemia, male pattern baldness, and  
 CC bacterial, fungal, protozoan and viral infections. The kringle1  
 CC polypeptides may also be used to generate antibodies. Determining the  
 CC presence or absence of mutations in, and analysing for the presence or  
 CC absence of expression of, kringle1 polynucleotides can be used to  
 CC diagnose a disease or susceptibility to a disease related to expression  
 CC or activity of kringle1 proteins. The polynucleotides may also be used  
 CC for chromosome identification, and mapping  
 XX  
 SQ Sequence 286 AA;  
 Query Match 19.4%; Score 47; DB 2; Length 286;  
 Best Local Similarity 100.0%; Pred. No. 1.2e-33;  
 Matches 47; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 45 NHSYCRNPDEPRGVCVSGAGVPEKRPCEDLRCPETTSQALPAF 91  
 DB 66 NHSYCRNPDEPRGVCVSGAGVPEKRPCEDLRCPETTSQALPAF 112  
 RESULT 19  
 AAY12397  
 ID AAY12397 standard; protein; 55 AA.  
 XX  
 AC AAY12397;  
 XX  
 DT 17-JUN-1999 (first entry)  
 XX

KW upstream regulatory sequence; cytokine activity; cell proliferation; differentiation; haematopoiesis regulation; tissue growth regulation;  
 KW reproductive hormone regulation; chemotactic; chemokinetic; haemostatic;  
 KW thrombolytic; antiinflammatory; tumour inhibition; antitumour.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO9906553-A2.  
 XX  
 PD 11-FEB-1999.  
 XX  
 PF 31-JUL-1998; 98WO-IB001237.  
 XX  
 PR 01-AUG-1997; 97US-00905051.  
 XX  
 PA (GEST ) GENSET.  
 XX  
 PI Dumas Milne Edwards J, Duclert A, Lacroix B;  
 XX  
 DR WPI; 1999-153783/13.  
 DR N-P8DB; AAX41473.  
 XX  
 PT New nucleic acids encoding human secreted proteins - obtained from cDNA  
 PT libraries derived from umbilical cord, lymph ganglia, lymphocytes and  
 PT placental tissue.  
 XX  
 PS Claim 34; Page 376; 411pp; English.  
 XX  
 CC AAX41379 to AAX41526 represent 5' expressed sequence tags (ESTs) for  
 CC human secreted proteins, and encode the proteins given in AAY12521 to  
 CC AAY12668, respectively. The proteins given represent the signal peptide  
 CC and an N-terminal fragment of a secreted protein. The nucleic acid  
 CC sequences can be used for producing secreted human gene products. They  
 CC can also be used to develop products for diagnosis and therapy. The  
 CC proteins obtained may have cytokine activity, cell  
 CC proliferation/differentiation activity, haematopoiesis regulating  
 CC activity, tissue growth regulating activity, reproductive hormone  
 CC regulating activity, chemotactic/chemokinetic activity, haemostatic and  
 CC thrombolytic activity, receptor/ligand activity, antiinflammatory  
 CC activity, tumour inhibition activity or other activities. The products  
 CC can be used in forensic, gene therapy and chromosome mapping procedures.  
 CC The sequences can also be used for obtaining corresponding promoter  
 CC sequences. The nucleic acids encoding the signal peptide can be used for  
 CC directing extracellular secretion of a polypeptide or the insertion of a  
 CC polypeptide into a membrane, or importing a polypeptide into a cell  
 XX  
 SQ Sequence 56 AA;  
 Query Match 13.2%; Score 32; DB 2; Length 56;  
 Best Local Similarity 100.0%; Pred. No. 8.1e-21;  
 Matches 32; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 OY 1 SGGCFWDNGHLYREDQTSAPGLRCLNWLDAQ 32  
 |||||  
 Db 22 SGGCFWDNGHLYREDQTSAPGLRCLNWLDAQ 53  
 |||||  
 RESULT 21  
 AAW72640  
 ID AAW72640 standard; peptide; 39 AA.  
 XX  
 AC AAW72640;  
 XX  
 DT 05-JAN-1999 (first entry)  
 XX  
 DE Nervous glia cell growth factor N-terminal peptide #1.  
 XX  
 DE Nervous glia cell growth factor; human; urine; secretion promoter;  
 KW choline acetyltransferase activity enhancer; nervous disease.  
 XX  
 OS Homo sapiens.  
 XX  
 FH Key Location/Qualifiers

FT Misc-difference 25 /note= "unspecified"  
 FT Misc-difference 29 /note= "unspecified"  
 FT  
 XX  
 PN JP10265498-A.  
 XX  
 XX 06-OCT-1998.  
 XX  
 XX 24-MAR-1997; 97JP-00090305.  
 XX  
 XX 24-MAR-1997; 97JP-00090305.  
 XX  
 PA (NICH-) JAPAN CHEM RES CO LTD.  
 XX  
 XX WPI; 1998-589719/50.  
 XX  
 XX Nervous glia cell growth factor derived from human urine - used for  
 PT treatment of nervous diseases.  
 PT  
 XX Claim 2; Fig 6; 14pp; Japanese.  
 XX  
 CC The present invention describes nervous glia cell growth factor, which is  
 CC purified from human urine by ultrafiltration, salting-out by ammonium  
 CC sulphate, gel filtration, ion exchange chromatography and reversed phase  
 CC chromatography, and has a M.W. of 29 kDa by sodium dodecyl sulphate-  
 CC polyacrylamide gel electrophoresis. Also described are: (1) a secretion  
 CC promoter for the nerve growth factor of glia cell consisting of the above  
 CC growth factor, an enhancer for choline acetyltransferase activity of  
 CC neuron consisting of the above growth factor; and (2) DNA encoding  
 CC nervous glia growth factor containing a DNA sequence coding the amino  
 CC acid sequence shown by the two 39 amino acid sequences as given in  
 CC AAW72640 and AAW72641, which are identical, except one starts with Tyr  
 CC and the other with Ser (i.e. they are from different DNA transcripts).  
 CC The glia cell growth factor can be prepared in a large amount and the  
 CC factor can be used for the treatment of nervous diseases  
 XX  
 SQ Sequence 39 AA;  
 Query Match 8.3%; Score 20; DB 2; Length 39;  
 Best Local Similarity 100.0%; Pred. No. 3.3e-10;  
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 OY 5 FWDNGHLYREDQTSAPGLR 24  
 |||||  
 Db 5 FWDNGHLYREDQTSAPGLR 24  
 |||||  
 RESULT 22  
 AAW72641  
 ID AAW72641 standard; peptide; 39 AA.  
 XX  
 AC AAW72641;  
 XX  
 DT 05-JAN-1999 (first entry)  
 XX  
 DE Nervous glia cell growth factor N-terminal peptide #2.  
 XX  
 DE Nervous glia cell growth factor; human; urine; secretion promoter;  
 KW choline acetyltransferase activity enhancer; nervous disease.  
 XX  
 OS Homo sapiens.  
 XX  
 FH Key Location/Qualifiers  
 FT Misc-difference 25 /note= "unspecified"  
 FT Misc-difference 29 /note= "unspecified"  
 FT  
 XX  
 PN JP10265498-A.  
 XX  
 XX 06-OCT-1998.  
 XX

PF 24-MAR-1997; 97JP-00090305.  
XX  
PR 24-MAR-1997; 97JP-00090305.  
XX  
PA (NICH-) JAPAN CHEM RES CO LTD.  
XX  
XX WPI; 1998-589719/50.  
XX  
PT Nervous glia cell growth factor derived from human urine - used for  
PT treatment of nervous diseases.  
XX  
PS Claim 3; Fig 7; 14pp; Japanese.  
XX  
XX The present invention describes nervous glia cell growth factor, which is  
CC purified from human urine by ultrafiltration, salting-out by ammonium  
CC sulphate, gel filtration, ion exchange chromatography and reversed-phase  
CC chromatography, and has a M.W. of 29 kDa by sodium dodecyl sulphate-  
CC polyacrylamide gel electrophoresis. Also described are: (1) a secretion  
CC promoter for the nerve growth factor of glia cell consisting of the above  
CC growth factor, an enhancer for choline acetyltransferase activity of  
CC neuron consisting of the above growth factor; and (2) DNA encoding  
CC nervous glia growth factor containing a DNA sequence coding the amino  
CC acid sequence shown by the two 39 amino acid sequences as given in  
CC AAW72640 and AAW72641, which are identical, except one starts with Tyr  
CC and the other with Ser (i.e. they are from different DNA transcripts).  
CC The glia cell growth factor can be prepared in a large amount and the  
CC factor can be used for the treatment of nervous diseases  
XX  
SQ Sequence 39 AA;  
Query Match 8.3%; Score 20; DB 2; Length 39;  
Best Local Similarity 100.0%; Pred. No. 3.3e-10;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 5 FWDNGHLYREDQTSAPGLR 24  
DB 5 FWDNGHLYREDQTSAPGLR 24  
RESULT 23  
ABR42624  
ID ABR42624 standard; protein; 81 AA.  
XX  
AC ABR42624;  
XX  
XX 26-AUG-2003 (first entry)  
DT  
DE Human kringle containing protein.  
XX  
XX Human; abrogen; kringle; angiogenesis; inhibitor; tumour; metastasis;  
XX cytostatic; gene therapy; expressed sequence tag; EST.  
XX  
XX Homo sapiens.  
XX  
XX Key Location/Qualifiers  
FT Misc-difference 4 /note= "the identity of this residue is unclear in Fig 2"  
FT Misc-difference 6 /note= "the identity of this residue is unclear in Fig 2"  
FT Misc-difference 7 /note= "the identity of this residue is unclear in Fig 2"  
FT Misc-difference 16 /note= "the identity of this residue is unclear in Fig 2"  
FT Misc-difference 29 /note= "the identity of this residue is unclear in Fig 2"  
FT Misc-difference 32 /note= "the identity of this residue is unclear in Fig 2"  
FT Misc-difference 46 /note= "the identity of this residue is unclear in Fig 2"  
FT Domain 50..56 /note= "kringle domain"  
FT Misc-difference 57 /note= "the identity of this residue is unclear in Fig 2"

FT Misc-difference 62 /note= "the identity of this residue is unclear in Fig 2"  
FT Misc-difference 70 /note= "the identity of this residue is unclear in Fig 2"  
FT Misc-difference 72 /note= "the identity of this residue is unclear in Fig 2"  
FT Misc-difference 77 /note= "the identity of this residue is unclear in Fig 2"  
XX WO2003042354-A2.  
XX  
XX 22-MAY-2003.  
XX  
XX 04-SEP-2002; 2002WO-US027885.  
XX  
XX 04-SEP-2001; 2001US-0316300P.  
XX (AVET ) AVENTIS PHARM INC.  
XX Nesbit M, Fong TC, Brockstedt D;  
XX WPI; 2003-449566/42.  
XX  
XX New abrogen polypeptide, useful for treating an angiogenesis related  
XX diseases e.g. tumor metastasis.  
XX  
XX Disclosure; Fig 2; 95pp; English.  
XX  
XX The present sequence is the protein sequence of a hypothetical kringle-  
XX containing protein encoded by an expressed sequence tag. The invention  
XX relates to novel abrogen polypeptides that are derived from kringle-  
XX containing proteins. The abrogens are potent inhibitors of endothelial  
XX proliferation and angiogenesis. They are capable of inhibiting or  
XX reducing cell proliferation induced by both basic fibroblast growth  
XX factor and vascular endothelial growth factor in a specific endothelial  
XX cell proliferation assay. Vectors that expressed abrogen polypeptides in  
XX vivo were shown to reduce tumour metastasis in 2 lung cancer models. The  
XX invention provides abrogen polypeptides and polynucleotides, and methods  
XX of using these to treat an angiogenesis-related disease or disorder, e.g.  
XX tumour metastasis (claimed)  
XX  
SQ Sequence 81 AA;  
Query Match 5.4%; Score 13; DB 7; Length 81;  
Best Local Similarity 100.0%; Pred. No. 0.0011;  
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 34 GLSAPVSGAGNH 46  
DB 33 GLSAPVSGAGNH 45  
RESULT 24  
AAW54157  
ID AAW54157 standard; protein; 527 AA.  
XX  
AC AAW54157;  
XX  
XX 20-JUL-1998 (first entry)  
DT  
DE t-PA mutant (N142S).  
XX  
XX Amino acid substitution; t-PA; vascular disorder; prevention;  
XX fibrin deposition; adhesion formation.  
XX Synthetic.  
XX  
XX Key Location/Qualifiers  
FT Domain 1..44 /note= "Finger domain"  
FT Domain 45..91 /note= "Growth factor domain"  
FT Domain 92..173

```

FT Misc-difference 142 /note= "Krangle-1 domain"
FT 142
FT Domain /note= "N changed from wt to S in mutant"
FT 180..261
FT Domain /note= "Krangle-2 domain"
FT 264..527
FT Domain /note= "Serine protease domain"
XX
XX US5736135-A.
XX
XX 07-APR-1998.
XX
XX 13-FEB-1995; 95US-00389615.
XX
XX 11-JUL-1991; 91US-00728456.
XX
XX 26-JAN-1993; 93US-00008940.
XX
XX 01-APR-1994; 94US-00221660.
XX
XX (GETH ) GENENTECH INC.
XX
XX Rice GC, Leung DW, Goeddel DV;
XX WPI; 1998-239153/21.
XX
XX Mutant tissue plasminogen activator proteins - useful for treating
XX vascular disorders, preventing tissue adhesion(s), etc.
XX
XX Claim 2; Page; 24pp; English.
XX
XX Mutant tissue plasminogen activator proteins (AAW54147-W54158) are
XX created by single or multiple amino acid substitutions. Compositions
XX containing the t-PA variant are used for treating vascular disorders, for
XX preventing fibrin deposition or for preventing adhesion formation or
XX reformation. Note: This sequence is not given in the specification but
XX was created from the wild type by the indexer
XX
XX Sequence 527 AA;
SQ
Query Match 4.1%; Score 10; DB 2; Length 527;
Best Local Similarity 100.0%; Pred. No. 2.7;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 44 GNHSYCRNPD 53
Db 139 GNHSYCRNPD 148

RESULT 25
AAW54154
ID AAW54154 standard; protein; 527 AA.
XX
XX AAW54154;
XX
XX 20-JUL-1998 (first entry)
XX
XX t-PA mutant (N142S).
XX
XX Amino acid substitution; t-PA; vascular disorder; prevention;
XX fibrin deposition; adhesion formation.
XX
XX Synthetic.
XX
XX Key Location/Qualifiers
XX Domain 1..44
XX /note= "Finger domain"
XX Domain 45..91
XX /note= "Growth factor domain"
XX Domain 92..173
XX /note= "Krangle-1 domain"
XX
XX Misc-difference 142 /note= "N changed from wt to S in mutant"
XX 180..261
XX Domain /note= "Krangle-2 domain"
XX

FT Misc-difference 142 /note= "Krangle-1 domain"
FT 142
FT Domain /note= "N changed from wt to S in mutant"
FT 180..261
FT Domain /note= "Krangle-2 domain"
FT 264..527
FT Domain /note= "Serine protease domain"
XX
XX US5736135-A.
XX
XX 07-APR-1998.
XX
XX 13-FEB-1995; 95US-00389615.
XX
XX 11-JUL-1991; 91US-00728456.
XX
XX 26-JAN-1993; 93US-00008940.
XX
XX 01-APR-1994; 94US-00221660.
XX
XX (GETH ) GENENTECH INC.
XX
XX Rice GC, Leung DW, Goeddel DV;
XX WPI; 1998-239153/21.
XX
XX Mutant tissue plasminogen activator proteins - useful for treating
XX vascular disorders, preventing tissue adhesion(s), etc.
XX
XX Claim 2; Page; 24pp; English.
XX
XX Mutant tissue plasminogen activator proteins (AAW54147-W54158) are
XX created by single or multiple amino acid substitutions. Compositions
XX containing the t-PA variant are used for treating vascular disorders, for
XX preventing fibrin deposition or for preventing adhesion formation or
XX reformation. Note: This sequence is not given in the specification but
XX was created from the wild type by the indexer
XX
XX Sequence 527 AA;
SQ
Query Match 4.1%; Score 10; DB 2; Length 527;
Best Local Similarity 100.0%; Pred. No. 2.7;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 44 GNHSYCRNPD 53
Db 139 GNHSYCRNPD 148

RESULT 26
ABU16473
ID ABU16473 standard; protein; 331 AA.
XX
XX ABU16473;
XX
XX 19-JUN-2003 (first entry)
XX
XX Protein encoded by Prokaryotic essential gene #2000.
XX
XX Antisense; prokaryotic essential gene; cell proliferation; drug design.
XX
XX Staphylococcus aureus.
XX
XX WO200277183-A2.
XX
XX 03-OCT-2002.
XX
XX 21-MAR-2002; 2002WO-US009107.
XX
XX 21-MAR-2001; 2001US-00815242.
XX
XX 06-SEP-2001; 2001US-00948993.
XX
XX 25-OCT-2001; 2001US-0342923P.
XX
XX 08-FEB-2002; 2002US-00072851.
XX
XX 06-MAR-2002; 2002US-0362699P.
XX
XX (ELIT-) ELITRA PHARM INC.
XX
XX Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;
XX Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;
XX

```



DR WPI; 2003-029926/02.  
DR N-PSDB; ACA20343.  
XX  
PT New antisense nucleic acids, useful for identifying proteins or screening  
PT for homologous nucleic acids required for cellular proliferation to  
PT isolate candidate molecules for rational drug discovery programs.  
XX  
PS Claim 25; SEQ ID NO 44397; 1766pp; English.  
XX  
XX The invention relates to an isolated nucleic acid comprising any one of  
CC the 6213 antisense sequences given in the specification where expression  
CC of the nucleic acid inhibits proliferation of a cell. Also included are:  
CC (1) a vector comprising a promoter operably linked to the nucleic acid  
CC encoding a polypeptide whose expression is inhibited by the antisense  
CC nucleic acid; (2) a host cell containing the vector; (3) an isolated  
CC polypeptide or its fragment whose expression is inhibited by the  
CC antisense nucleic acid; (4) an antibody capable of specifically binding  
CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular  
CC proliferation or the activity of a gene in an operon required for  
CC proliferation; (7) identifying a compound that influences the activity of  
CC the gene product or that has an activity against a biological pathway  
CC required for proliferation, or that inhibits cellular proliferation; (8)  
CC identifying a gene required for cellular proliferation or the biological  
CC pathway in which a proliferation-required gene or its gene product lies  
CC or a gene on which the test compound that inhibits proliferation of an  
CC organism acts; (9) manufacturing an antibiotic; (10) profiling a  
CC compound's activity; (11) a culture comprising strains in which the gene  
CC product is overexpressed or underexpressed; (12) determining the extent  
CC to which each of the strains is present in a culture or collection of  
CC strains; or (13) identifying the target of a compound that inhibits the  
CC proliferation of an organism. The antisense nucleic acids are useful for  
CC identifying proteins or screening for homologous nucleic acids required  
CC for cellular proliferation to isolate candidate molecules for rational  
CC drug discovery programs, or for screening homologous nucleic acids  
CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,  
CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is encoded by one of  
CC the target prokaryotic essential genes. Note: The sequence data for this  
CC patent did not form part of the printed specification, but was obtained  
CC in electronic format directly from WIFO at  
CC ftp.wipo.int/pub/published\_pct\_sequences  
XX  
SQ Sequence 331 AA;

Query Match 3.7%; Score 9; DB 6; Length 331;  
Best Local Similarity 100.0%; Pred. No. 14;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 160 IIAIGAGII 168  
| | | | | | | | | |  
Db 97 IIAIGAGII 105  
  
RESULT 27  
ABM72983  
ID ABM72983 standard; protein; 357 AA.  
XX  
AC ABM72983;  
XX  
DT 20-NOV-2003 (first entry)  
XX  
DE Staphylococcus aureus protein #2223.  
XX  
KW Antibacterial; vaccine; gene therapy; infection; sepsis; diagnosis;  
KW enzymatic assay; antibiotic target.  
XX  
OS Staphylococcus aureus.  
XX  
FN WO200294868-A2.  
XX  
PD 28-NOV-2002.  
XX  
PF 27-MAR-2002; 2002WO-IB002637.  
XX

PR 27-MAR-2001; 2001GB-00007661.  
PA (CHIR-) CHIRON SPA.  
XX  
XX Masignani V, Mora M, Scarselli M;  
PI WPI; 2003-120786/11.  
XX  
DR N-PSDB; ACP74543.  
XX  
XX New Staphylococcus aureus protein, useful as a vaccine for treating or  
PT preventing Staphylococcal infection, specifically an infection caused by  
PT *S. aureus*, e.g. sepsis.  
XX  
PS Claim 1; SEQ ID NO 4446; 49pp; English.  
XX  
XX The invention relates to novel genes and encoded proteins from  
CC Staphylococcus aureus. A composition comprising the *S. aureus* protein, a  
CC nucleic acid encoding the protein, or an antibody to the protein, is  
CC useful as a pharmaceutical, particularly as a vaccine for treating or  
CC preventing infection due to Staphylococcus bacteria, specifically an  
CC infection caused by *S. aureus*. The composition is particularly useful for  
CC treating or preventing sepsis in a patient. The composition can also be  
CC used for diagnostics. The protein is also used in an assay for enzymatic  
CC studies and as a target for antibiotics. This sequence represents one of  
CC the novel *S. aureus* proteins of the invention  
XX  
SQ Sequence 357 AA;  
  
Query Match 3.7%; Score 9; DB 6; Length 357;  
Best Local Similarity 100.0%; Pred. No. 15;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 160 IIAIGAGII 168  
| | | | | | | | | |  
Db 123 IIAIGAGII 131  
  
RESULT 28  
ABM69063  
ID ABM69063 standard; protein; 514 AA.  
XX  
AC ABM69063;  
XX  
XX 20-NOV-2003 (first entry)  
XX  
DE Photorhabdus luminescens protein sequence #2160.  
XX  
KW Antibacterial; fungicide; insecticide; polymorphism; genetic analysis;  
KW detection; food; gene expression; plant; animal; microorganism; toxin;  
KW antibiotic; biopesticide; virulence factor; disease model; plague;  
KW whooping cough.  
XX  
OS Photorhabdus luminescens.  
XX  
FN WO200294867-A2.  
XX  
PD 28-NOV-2002.  
XX  
PF 07-FEB-2002; 2002WO-IB003040.  
XX  
PR 07-FEB-2001; 2001FR-00001659.  
XX  
XX (INSP) INST PASTEUR.  
PA (CNRS) CNRS CENT NAT RECH SCI.  
XX  
XX Duchaud E, Taourit S, Glaser P, Frangeul L, Kunst F, Danchin A;  
PI Buchrieser C;  
XX  
XX WPI; 2003-148459/14.  
XX  
XX Genomic sequence of Photorhabdus luminescens and encoded polypeptides,  
PT useful e.g. as therapeutic antimicrobials and agricultural pesticides.  
XX

PS Claim 2; SEQ ID NO 2160; 1205pp; French.

XX The invention relates to the isolation of genes and their encoded

CC proteins from Photobacterium luminescens. The isolated sequences are

CC sources of probes and primers for detecting the genome of P. luminescens

CC and related species; to study polymorphisms; for gene analysis and for

CC detection/amplification of the genes. Antibodies (Ab) raised against the

CC polypeptides encoded by the genes are used for detection/identification

CC of P. luminescens, e.g. in foods. The genes, proteins, Ab and cells that

CC carry a gene-containing vector are used to select compounds that

CC modulate, regulate, induce or inhibit expression of the genes in plants,

CC animals or microorganisms other than P. luminescens and are able to alter

CC response or sensitivity to toxins and antibiotics produced by P.

CC luminescens. Cells transformed to express the genes are useful for

CC recombinant production of the proteins, particularly toxins and

CC antibacterials useful as insecticides, bactericides and fungicides. The

CC genes, proteins, vectors containing the genes and Ab are also useful

CC therapeutically (to treat microbial infection by bacteria or fungi that

CC are sensitive to P. luminescens-encoded toxins or antibiotics) and as

CC biopesticides. Other uses of the genes and the proteins are as virulence

CC factors and for identifying targets of human diseases for which P.

CC luminescens is a model (particularly plague and whooping cough). This

CC sequence represents one of the isolated P. luminescens proteins

XX

SQ Sequence 514 AA;

Query Match 3.3%; Score 8; DB 6; Length 514;

Best Local Similarity 100.0%; Pred. No. 1.6e+02;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 33 SGLASAPV 40

DB 305 SGLASAPV 312

|||||||

RESULT 29

ABU23579

ID ABU23579 standard; protein; 602 AA.

XX

AC ABU23579;

DT 19-JUN-2003 (first entry)

XX

DE Protein encoded by Prokaryotic essential gene #9106.

XX

KW Antisense; prokaryotic essential gene; cell proliferation; drug design.

XX

OS Clostridium acetobutylicum.

XX

PN WO20027183-A2.

XX

PD 03-OCT-2002.

XX

PF 21-MAR-2002; 2002WO-US009107.

XX

PR 21-MAR-2001; 2001US-00815242.

PR 06-SEP-2001; 2001US-00948993.

PR 25-OCT-2001; 2001US-0342923P.

PR 08-FEB-2002; 2002US-00072851.

PR 06-MAR-2002; 2002US-0362699P.

XX

PA (ELIT-) ELITRA PHARM INC.

XX

XX Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;

PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;

XX

DR WPI; 2003-029926/02.

XX

DR N-PSDB; ACA27449.

XX

XX New antisense nucleic acids, useful for identifying proteins or screening

PT for homologous nucleic acids required for cellular proliferation to

PT isolate candidate molecules for rational drug discovery programs.

XX

PS Claim 25; SEQ ID NO 51503; 1766pp; English.

XX The invention relates to an isolated nucleic acid comprising any one of

CC the 6213 antisense sequences given in the specification where expression

CC of the nucleic acid inhibits proliferation of a cell. Also included are:

CC (1) a vector comprising a promoter operably linked to the nucleic acid

CC encoding a polypeptide whose expression is inhibited by the antisense

CC nucleic acid; (2) a host cell containing the vector; (3) an isolated

CC polypeptide or its fragment whose expression is inhibited by the

CC antisense nucleic acid; (4) an antibody capable of specifically binding

CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular

CC proliferation or the activity of a gene in an operon required for

CC proliferation; (7) identifying a compound that influences the activity of

CC the gene product or that has an activity against a biological pathway

CC required for proliferation, or that inhibits cellular proliferation; (8)

CC identifying a gene required for cellular proliferation or the biological

CC pathway in which a proliferation-required gene or its gene product lies

CC or a gene on which the test compound that inhibits proliferation of an

CC organism acts; (9) manufacturing an antibiotic; (10) profiling a

CC compound's activity; (11) a culture comprising strains in which the gene

CC product is overexpressed or underexpressed; (12) determining the extent

CC to which each of the strains is present in a culture or collection of

CC strains; or (13) identifying the target of a compound that inhibits the

CC proliferation of an organism. The antisense nucleic acids are useful for

CC identifying proteins or screening for homologous nucleic acids required

CC for cellular proliferation to isolate candidate molecules for rational

CC drug discovery programs, or for screening homologous nucleic acids

CC required for proliferation in cells other than S. aureus, S. typhimurium,

CC K. pneumoniae or P. aeruginosa. The present sequence is encoded by one of

CC the target prokaryotic essential genes. Note: The sequence data for this

CC patent did not form part of the printed specification, but was obtained

CC in electronic format directly from WIPO at

CC ftp.wipo.int/pub/published\_pct\_sequences

XX

SQ Sequence 602 AA;

Query Match 3.3%; Score 8; DB 6; Length 602;

Best Local Similarity 100.0%; Pred. No. 1.9e+02;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 63 VSGEAGVP 70

DB 213 VSGEAGVP 220

|||||||

RESULT 30

AAR66600

ID AAR66600 standard; protein; 701 AA.

XX

AC AAR66600;

XX

DT 13-FEB-1995 (first entry)

XX

XX Mouse L5/3 tumour suppressor protein.

XX

KW Mouse L5/3 gene; small lung cell carcinoma; tumour suppression;

KW chromosome 3; 3p21; D3F15S2 locus; hepatocyte growth factor;

KW renal cell carcinoma; Von Hippel-Lindau syndrome; predisposition;

XX kringie domain.

XX

OS Mus musculus.

XX

XX

PH Key Location/Qualifiers

FT Peptide 1..16

FT /label= signal\_peptide\_(16-31)

FT /note= "putative"

FT

FT Misc-difference 4

FT /label= polymorphic site

FT /note= "Pro corresponds to CCG codon in cDNA; in the

FT genomic DNA, codon 19 is CAG (Gln)"

FT

FT Protein 17..701

FT /note= "putative protein contains 4 kringie domains

FT followed by a serine protease-like domain"

FT

```

FT Modified-site 57..59 /label= N-glycosylation_site
FT /note= "potential"
FT Modified-site 158..160 /label= N-glycosylation_site
FT /note= "potential"
FT Modified-site 220..292 /label= N-glycosylation_site
FT /note= "potential"
FT Modified-site 605..607 /label= N-glycosylation_site
FT /note= "potential"
XX
XX
XX US5315000-A.
XX
XX 24-MAY-1994.
XX
XX 14-MAY-1992; 92US-00882925.
XX
XX 14-MAY-1992; 92US-00882925.
XX
XX (CHIL-) CHILDRENS HOSPITAL MEDICAL CENT.
XX
XX Degen SJ;
XX
XX WPI; 1994-166645/20.
XX N-PSDB; AAQ79726.
XX
XX DNA from D3F15S2 locus of human chromosome 3 - encoding hepatocyte
XX growth factor, L5/3, useful as probe for detecting pre-deposition towards
XX cancer.
XX
XX Disclosure; Col 27-32; 31pp; English.
XX
XX A mouse liver cDNA library was screened with cDNA coding for human L5/3;
XX the L5/3 gene, located at the D3F15S2 locus of human chromosome 3, codes
XX for a protein composed of 4 kringle domains, followed by a serine
XX protease-like domain. The longest murine clone to be isolated (pML5-2,
XX AAQ79726) was not full-length. The open reading frame was present at the
XX 5' end of the sequence with no codon for the initiator methionine in-frame
XX with the coding sequence. After determination of the sequence of the
XX mouse gene it was determined that the cDNA lacked 44bp of coding and 94bp
XX of non-coding sequence at its 5'-end
XX
XX Sequence 701 AA;
XX
Query Match 3.3%; Score 8; DB 2; Length 701;
Best Local Similarity 100.0%; Pred. No. 2.1e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 55 DPGPWCY 62
DB 148 DPGPWCY 155

RESULT 31
AAW14271
ID AAW14271 standard; protein; 701 AA.
XX
XX AAW14271;
XX
XX 25-MAR-2003 (revised)
XX 21-JUL-1997 (first entry)
XX
XX Mouse growth factor L5/3 partial cDNA clone ML5-2 encoded protein.
XX
XX Mouse; growth factor; foetal; liver; probe; bovine; prothrombin; locus;
XX polymorphism; transition; exon; intron; chromosome; kringle domain;
XX cell growth; tumour suppressor; hepatocyte growth factor; regeneration.
XX
XX Mus musculus.
XX
XX Key Location/Qualifiers

```

```

FT Peptide 1..16 /note= "partial signal peptide sequence"
FT Protein 17..701 /note= "mature protein"
FT Misc-difference 19 /note= "amino acid residue is Gln in this position in the
FT protein encoded by the genomic sequence (AAT62442); this
FT may be due to a polymorphism"
XX
XX US5606029-A.
XX
XX 25-FEB-1997.
XX
XX 18-JAN-1994; 94US-00184012.
XX
XX 14-MAY-1992; 92US-00882925.
XX
XX (CHIL-) CHILDREN'S HOSPITAL MEDICAL CENT.
XX
XX Degen SJ;
XX
XX WPI; 1997-153621/14.
XX N-PSDB; AAT62441.
XX
XX Human growth factor protein L5/3 - useful for altering cell growth, e.g.
XX as tumour suppressor.
XX
XX Disclosure; Col 27-32; 34pp; English.
XX
XX This is the amino acid sequence encoded by the insert isolated from clone
XX ML5-2 and constitutes part of a mouse growth factor designated L5/3
XX (AAW14272). The encoding sequence was isolated from a lambda-gt10 mouse
XX liver cDNA library using a fragment of the corresponding human cDNA
XX sequence (AAT62436) as a probe. The fragment presented here is a partial
XX sequence which lacks 44 bp of the coding region and 94 bp of the non-
XX coding region. A 1450 bp fragment of this sequence (nucleotides 738-2188)
XX covering 8 amino acids of the second kringle domain and all of the third
XX and fourth kringle domains and the serine protease-like domain, was used
XX to screen a mouse liver genomic DNA library to isolate the full length
XX genomic sequence (AAT62442). The protein can be used to alter cell growth
XX (as a growth factor or tumour suppressor) and has similar properties to
XX the hepatocyte growth factor that is actively involved in liver
XX regeneration. (Updated on 25-MAR-2003 to correct PF field.)
XX
XX Sequence 701 AA;
XX
Query Match 3.3%; Score 8; DB 2; Length 701;
Best Local Similarity 100.0%; Pred. No. 2.1e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 55 DPGPWCY 62
DB 148 DPGPWCY 155

RESULT 32
AAW66601
ID AAW66601 standard; protein; 716 AA.
XX
XX AAW66601;
XX
XX 14-FEB-1995 (first entry)
XX
XX Mouse L5/3 tumour suppressor protein (from genomic sequence).
XX
XX Mouse L5/3 gene; small lung cell carcinoma; tumour suppression;
XX chromosome 3; 3p21; D3F15S2 locus; hepatocyte growth factor;
XX renal cell carcinoma; Von Hippel-Lindau syndrome; predisposition;
XX kringle domain.
XX
XX Mus musculus.
XX
XX Key Location/Qualifiers

```

FT Peptide 1. .31  
FT /label= signal peptide  
FT /note= "putative"  
FT Misc-difference 19  
FT /label= polymorphic\_site  
FT /note= "Gln corresponds to CAG codon in genomic DNA; in  
FT the cDNA, codon 19 is CCG (pro) (AAR66600)"  
FT Protein 32. .716  
FT /note= "putative protein contains 4 kringle domains  
FT followed by a serine protease-like domain"  
FT Modified-site 72. .74  
FT /label= N-glycosylation\_site  
FT /note= "potential"  
FT Modified-site 173. .175  
FT /label= N-glycosylation\_site  
FT /note= "potential"  
FT Modified-site 305. .307  
FT /label= N-glycosylation\_site  
FT /note= "potential"  
FT Modified-site 620. .622  
FT /label= N-glycosylation\_site  
FT /note= "potential"  
FT US5315000-A.  
FT 24-MAY-1994.  
FT 14-MAY-1992; 92US-00882925.  
FT 14-MAY-1992; 92US-00882925.  
FT (CHIL-) CHILDRENS HOSPITAL MEDICAL CENT.  
FT Degen SJ;  
FT WPI; 1994-166645/20.  
FT N-PSDB; AAQ79727.  
FT DNA from D3F15S2 locus of human chromosome 3- - encoding hepatocyte  
FT growth factor, L5/3, useful as probe for detecting pre-deposition towards  
FT cancer.  
FT Disclosure; Col 33-42; 3lpp; English.  
FT The sequence of mouse genomic DNA coding for the L5/3 tumour suppressor  
FT protein is composed of 18 exons separated by 17 intervening sequences.  
FT There is only one difference found between the cDNA (AAQ79726) and  
FT genomic DNA (AAQ79727) coding sequences which results in the substitution  
FT of a Gln in the gene to a Pro in the cDNA at amino acid position 19. The  
FT putative mouse protein has the same domain structure as its human  
FT homologue with four kringle domains followed by a serine protease-like  
FT domain  
FT  
FT Sequence 716 AA;  
FT  
FT Query Match 3.3%; Score 8; DB 2; Length 716;  
FT Best Local Similarity 100.0%; Pred. No. 2.2e+02;  
FT Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
FT  
FT QY 55 DPRGPWCY 62  
FT 163 DPRGPWCY 170  
FT  
FT Db  
FT  
FT RESULT 33  
FT AA14272  
FT ID AA14272 standard; protein; 716 AA.  
FT XX AA14272;  
FT AC AA14272;  
FT XX  
FT 25-MAR-2003 (revised)  
FT DT 21-JUL-1997 (first entry)  
FT XX

DE Mouse growth factor L5/3 complete protein.  
XX  
XX Mouse; growth factor; foetal; liver; probe; bovine; prothrombin; locus;  
KW polymorphism; transition; exon; intron; chromosome; kringle domain;  
KW Cell growth; tumour suppressor; hepatocyte growth factor; regeneration.  
XX  
OS Mus musculus.  
XX  
FH Key Location/Qualifiers  
FT Peptide 1. .31  
FT /note= "signal peptide"  
FT Misc-difference 19  
FT /note= "amino acid residue is pro at this position in the  
FT protein encoded by the cDNA clone ML5-2 (AA62441); this  
FT may be due to a polymorphism"  
FT Protein 32. .716  
FT /note= "mature protein"  
FT Modified-site 72  
FT /note= "N-linked glycosylation site"  
FT Modified-site 173  
FT /note= "N-linked glycosylation site"  
FT Modified-site 305  
FT /note= "N-linked glycosylation site"  
FT Modified-site 624  
FT /note= "N-linked glycosylation site"  
FT  
XX US5606029-A.  
XX PN  
XX 25-FEB-1997.  
XX  
XX 18-JAN-1994; 94US-00184012.  
XX 14-MAY-1992; 92US-00882925.  
XX  
XX (CHIL-) CHILDREN'S HOSPITAL MEDICAL CENT.  
XX Degen SJ;  
XX WPI; 1997-153621/14.  
XX N-PSDB; AA62442.  
XX Human growth factor protein L5/3 - useful for altering cell growth, e.g.  
XX as tumour suppressor.  
XX Disclosure; Col 33-42; 34pp; English.  
XX  
XX This is the amino acid sequence of the wild type mouse growth factor  
XX designated L5/3. The protein sequence differs from that encoded by the  
XX cDNA clone (AA62441) at position 19; in this sequence a Gln, in the cDNA  
XX clone a Pro. This difference may be due to a polymorphism at this codon.  
XX The full length mouse gene contains 18 exons and encodes a protein having  
XX a molecular weight 80 kD. The protein can be used to alter cell growth  
XX (as a growth factor or tumour suppressor) and has similar properties to  
XX the hepatocyte growth factor that is actively involved in liver  
XX regeneration. (Updated on 25-MAR-2003 to correct PF field.)  
XX  
XX Sequence 716 AA;  
XX  
XX Query Match 3.3%; Score 8; DB 2; Length 716;  
XX Best Local Similarity 100.0%; Pred. No. 2.2e+02;  
XX Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
XX  
XX QY 55 DPRGPWCY 62  
XX 163 DPRGPWCY 170  
XX  
XX Db  
XX  
XX RESULT 34  
XX AAY31156  
XX ID AAY31156 standard; protein; 716 AA.  
XX XX  
XX AC AAY31156;  
XX XX

DT 26-OCT-1999 (first entry)  
 XX Murine macrophage stimulating protein.  
 XX  
 XX  
 XX Macrophage stimulating protein; MSP; murine; modulator; proliferation;  
 KW differentiation; intestinal epithelium; colon crypt; treatment; cancer;  
 KW haematopoietic disorder; megakaryocyte deficiency; gastrointestinal;  
 KW chemotherapeutic agent; gut toxicity.  
 XX  
 XX Mus sp.  
 OS  
 XX US5948892-A.  
 PN  
 XX 07-SEP-1999.  
 PD  
 XX 16-DEC-1996; 96US-00766982.  
 PF  
 XX 16-DEC-1996; 96US-00766982.  
 PR  
 XX (AMGE-) AMGEN INC.  
 PA  
 XX Wahl RC;  
 XX  
 XX WPI; 1999-517975/43.  
 DR  
 XX Analogues of macrophage stimulating protein for treating gastrointestinal.  
 XX or haematopoietic disorders.  
 PT  
 XX  
 XX Claim 4; Col 15-18; 23pp; English.  
 PS  
 XX This invention describes a novel purified and isolated analogue of mature  
 CC macrophage stimulating protein (MSP) having at least one unpaired  
 CC cysteine residue substituted with another amino acid which modulates the  
 CC proliferation or differentiation of the intestinal epithelium. The  
 CC product of the invention binds to RON (a cell membrane protein tyrosine  
 CC kinase which is a member of the c-met family) to promote the formation of  
 CC colon crypts. MSP analogues are useful for the treatment of conditions  
 CC requiring the administration of MSP, such conditions include  
 CC haematopoietic disorders such as those involving a deficiency of  
 CC megakaryocytes and gastrointestinal disorders such as ulcerative colitis,  
 CC Crohn's disease and infections. The MSP analogues are useful for  
 CC maintaining and repairing the epithelial lining in the treatment of  
 CC cancer, where the aggressive use of chemotherapeutic agents or the use of  
 CC whole body radiation may lead to gut toxicity. The MSP analogues, which  
 CC have a higher activity than normal human MSP are effective at smaller  
 CC dosages, or optionally, they may be administered less frequently than  
 CC human MSP. This sequence represents the murine MSP described in the  
 CC method of the invention  
 XX  
 XX Sequence 716 AA;

Query Match 3.3%; Score 8; DB 2; Length 716;  
 Best Local Similarity 100.0%; Pred. No. 2.2e+02;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 55 DPRGPWCY 62  
 |||||  
 Db 163 DPRGPWCY 170

RESULT 35  
 AAW82790  
 ID AAW82790 standard; protein; 716 AA.  
 XX  
 XX AAW82790;

DT 12-APR-1999 (first entry)

XX Mouse MSP protein.

DE MSP; macrophage stimulating protein; apoptosis; murine; treatment;  
 KW neuroendocrine cell; RON receptor; small cell lung carcinoma; tumour;  
 KW pathogen infection; thrombocyte production; megakaryocyte maturation;

KW thrombocytopaenia; hepatocyte growth.  
 XX  
 OS Mus sp.  
 PN W09855141-A1.  
 XX  
 XX 10-DEC-1998.  
 PD  
 XX 04-JUN-1998; 98WO-US011573.  
 PF  
 XX 04-JUN-1997; 97US-0048594P.  
 PR  
 XX (BGHM ) BRIGHAM & WOMENS HOSPITAL.  
 FA  
 XX Sunday ME, Willet C;  
 XX WPI; 1999-059877/05.  
 XX N-PSDB; AAV72084.  
 DR  
 XX Treating tumours derived from neuroendocrine cells with macrophage  
 PT stimulating protein - or its nucleic acid, also for preventing  
 PT development of these tumours, specifically small cell lung carcinoma.  
 XX  
 PS Disclosure; Page 67-69; 100pp; English.

XX This sequence represents a novel murine macrophage stimulating protein,  
 CC MSP, which is used in a method for the prophylactic treatment of a tumour  
 CC derived from neuroendocrine cells (NEC) by administration of this MSP to  
 CC a subject at risk, sufficient to induce apoptosis of NEC expressing a RON  
 CC receptor (the receptor for MSP). The method is especially used to treat  
 CC or prevent small cell lung carcinoma and apoptosis of RON-expressing  
 CC cells may be induced in vivo or in vitro. Screening NEC from a subject  
 CC for susceptibility to MSP-induced apoptosis is used to identify patients  
 CC who will benefit from treatment with the MSP protein. MSP is already  
 CC known for treating pathogen infections, for stimulating thrombocyte  
 CC production and megakaryocyte maturation (for treating thrombocytopaenia)  
 CC and for stimulating growth of cells (particularly hepatocytes)

XX Sequence 716 AA;

Query Match 3.3%; Score 8; DB 2; Length 716;  
 Best Local Similarity 100.0%; Pred. No. 2.2e+02;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 55 DPRGPWCY 62  
 |||||  
 Db 163 DPRGPWCY 170

RESULT 36  
 ABU34603  
 ID ABU34603 standard; protein; 771 AA.  
 XX  
 XX AC ABU34603;

DT 19-JUN-2003 (first entry)

XX Protein encoded by Prokaryotic essential gene #20130.

DE Antisense; prokaryotic essential gene; cell proliferation; drug design.

KW Mycobacterium bovis.

XX W0200277183-A2.

XX 03-OCT-2002.

XX 21-MAR-2002; 2002WO-US009107.

XX 21-MAR-2001; 2001US-00815242.

PR 06-SEP-2001; 2001US-00948993.

PR 25-OCT-2001; 2001US-0342923P.

PR 08-FEB-2002; 2002US-00072851.

```
PR 06-MAR-2002; 2002US-0362699P.
XX (ELIT-) ELITRA PHARM INC.
XX
XX Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;
XX Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;
XX WPI; 2003-029926/02.
XX N-PSDB; ACA38473.
XX
XX New antisense nucleic acids, useful for identifying proteins or screening
XX for homologous nucleic acids required for cellular proliferation to
XX isolate candidate molecules for rational drug discovery programs.
XX
XX Claim 25; SEQ ID NO 62527; 1766pp; English.
XX
XX The invention relates to an isolated nucleic acid comprising any one of
XX the 6213 antisense sequences given in the specification where expression
XX of the nucleic acid inhibits proliferation of a cell. Also included are:
XX (1) a vector comprising a promoter operably linked to the nucleic acid
XX encoding a polypeptide whose expression is inhibited by the antisense
XX nucleic acid; (2) a host cell containing the vector; (3) an isolated
XX antisense nucleic acid; (4) an antibody capable of specifically binding
XX the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular
XX proliferation or the activity of a gene in an operon required for
XX proliferation; (7) identifying a compound that influences the activity of
XX the gene product or that has an activity against a biological pathway
XX required for proliferation, or that inhibits cellular proliferation; (8)
XX identifying a gene required for cellular proliferation or the biological
XX pathway in which a proliferation-required gene or its gene product lies
XX or a gene on which the test compound that inhibits proliferation of an
XX organism acts; (9) manufacturing an antibiotic; (10) profiling a
XX compound's activity; (11) a culture comprising strains in which the gene
XX product is overexpressed or underexpressed; (12) determining the extent
XX to which each of the strains is present in a culture or collection of
XX strains; or (13) identifying the target of a compound that inhibits the
XX proliferation of an organism. The antisense nucleic acids are useful for
XX identifying proteins or screening for homologous nucleic acids required
XX for cellular proliferation to isolate candidate molecules for rational
XX drug discovery programs, or for screening homologous nucleic acids
XX required for proliferation in cells other than S. aureus, S. typhimurium,
XX K. pneumoniae or P. aeruginosa. The present sequence is encoded by one of
XX the target prokaryotic essential genes. Note: The sequence data for this
XX patent did not form part of the printed specification, but was obtained
XX in electronic format directly from WIPO at
XX ftp.wipo.int/pub/published_pct_sequences
XX
XX Sequence 771 AA;
XX
XX Query Match 3.3%; Score 8; DB 6; Length 771;
XX Best Local Similarity 100.0%; Pred. No. 2.3e+02;
XX Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX QY 37 SAPVSGAG 44
XX |||||
XX 692 SAPVSGAG 699
XX
XX RESULT 37
XX ABU36546
XX ID ABU36546 standard; protein; 771 AA.
XX XX
XX AC ABU36546;
XX
XX XX 19-JUN-2003 (first entry)
XX
XX XX Protein encoded by Prokaryotic essential gene #22073.
XX
XX DE Antisense; prokaryotic essential gene; cell proliferation; drug design.
XX
XX KW Mycobacterium tuberculosis.
XX
XX OS
XX XX
```

```
PN WO200277183-A2.
XX 03-OCT-2002.
XX
XX 21-MAR-2002; 2002WO-US009107.
XX
XX 21-MAR-2001; 2001US-00815242.
XX 06-SEP-2001; 2001US-00948993.
XX 25-OCT-2001; 2001US-0342923P.
XX 08-FEB-2002; 2002US-00072851.
XX 06-MAR-2002; 2002US-0362699P.
XX
XX (ELIT-) ELITRA PHARM INC.
XX
XX Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;
XX Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;
XX WPI; 2003-029926/02.
XX N-PSDB; ACA40416.
XX
XX New antisense nucleic acids, useful for identifying proteins or screening
XX for homologous nucleic acids required for cellular proliferation to
XX isolate candidate molecules for rational drug discovery programs.
XX
XX Claim 25; SEQ ID NO 64470; 1766pp; English.
XX
XX The invention relates to an isolated nucleic acid comprising any one of
XX the 6213 antisense sequences given in the specification where expression
XX of the nucleic acid inhibits proliferation of a cell. Also included are:
XX (1) a vector comprising a promoter operably linked to the nucleic acid
XX encoding a polypeptide whose expression is inhibited by the antisense
XX nucleic acid; (2) a host cell containing the vector; (3) an isolated
XX polypeptide or its fragment whose expression is inhibited by the
XX antisense nucleic acid; (4) an antibody capable of specifically binding
XX the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular
XX proliferation or the activity of a gene in an operon required for
XX proliferation; (7) identifying a compound that influences the activity of
XX the gene product or that has an activity against a biological pathway
XX required for proliferation, or that inhibits cellular proliferation; (8)
XX identifying a gene required for cellular proliferation or the biological
XX pathway in which a proliferation-required gene or its gene product lies
XX or a gene on which the test compound that inhibits proliferation of an
XX organism acts; (9) manufacturing an antibiotic; (10) profiling a
XX compound's activity; (11) a culture comprising strains in which the gene
XX product is overexpressed or underexpressed; (12) determining the extent
XX to which each of the strains is present in a culture or collection of
XX strains; or (13) identifying the target of a compound that inhibits the
XX proliferation of an organism. The antisense nucleic acids are useful for
XX identifying proteins or screening for homologous nucleic acids required
XX for cellular proliferation to isolate candidate molecules for rational
XX drug discovery programs, or for screening homologous nucleic acids
XX required for proliferation in cells other than S. aureus, S. typhimurium,
XX K. pneumoniae or P. aeruginosa. The present sequence is encoded by one of
XX the target prokaryotic essential genes. Note: The sequence data for this
XX patent did not form part of the printed specification, but was obtained
XX in electronic format directly from WIPO at
XX ftp.wipo.int/pub/published_pct_sequences
XX
XX Sequence 771 AA;
XX
XX Query Match 3.3%; Score 8; DB 6; Length 771;
XX Best Local Similarity 100.0%; Pred. No. 2.3e+02;
XX Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX QY 37 SAPVSGAG 44
XX |||||
XX 692 SAPVSGAG 699
XX
XX Db
XX
XX RESULT 38
XX ABU34316
XX ID ABU34316 standard; protein; 774 AA.
XX
XX XX
```

```
AC ABU34316;
XX
XX 19-JUN-2003 (first entry)
XX
XX Protein encoded by Prokaryotic essential gene #19843.
XX
XX Antisense; prokaryotic essential gene; cell proliferation; drug design.
XX
XX Mycobacterium avium.
XX
XX WO200277183-A2.
XX
XX 03-OCT-2002.
XX
XX 21-MAR-2002; 2002WO-US009107.
XX
XX 21-MAR-2001; 2001US-00815242.
XX
XX 06-SEP-2001; 2001US-00948993.
XX
XX 25-OCT-2001; 2001US-0342923P.
XX
XX 08-FEB-2002; 2002US-00072851.
XX
XX 06-MAR-2002; 2002US-0362699P.
XX
XX (ELIT-) ELITRA PHARM INC.
XX
XX Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;
XX Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;
XX WPI; 2003-029926/02.
XX N-PSDB; ACA38186.
XX
XX New antisense nucleic acids, useful for identifying proteins or screening
XX PT for homologous nucleic acids required for cellular proliferation to
XX PT isolate candidate molecules for rational drug discovery programs.
XX
XX Claim 25; SEQ ID NO 62240; 1766pp; English.
XX
XX The invention relates to an isolated nucleic acid comprising any one of
XX the 6213 antisense sequences given in the specification where expression
XX of the nucleic acid inhibits proliferation of a cell. Also included are:
XX (1) a vector comprising a promoter operably linked to the nucleic acid
XX encoding a polypeptide whose expression is inhibited by the antisense
XX nucleic acid; (2) a host cell containing the vector; (3) an isolated
XX polypeptide or its fragment whose expression is inhibited by the
XX antisense nucleic acid; (4) an antibody capable of specifically binding
XX the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular
XX proliferation or the activity of a gene in an operon required for
XX proliferation; (7) identifying a compound that influences the activity of
XX the gene product or that has an activity against a biological pathway
XX required for proliferation, or that inhibits cellular proliferation; (8)
XX identifying a gene required for cellular proliferation or the biological
XX pathway in which a proliferation-required gene or its gene product lies
XX or a gene on which the test compound that inhibits proliferation of an
XX organism acts; (9) manufacturing an antibiotic; (10) profiling a
XX compound's activity; (11) a culture comprising strains in which the gene
XX product is overexpressed or underexpressed; (12) determining the extent
XX to which each of the strains is present in a culture or collection of
XX strains; or (13) identifying the target of a compound that inhibits the
XX proliferation of an organism. The antisense nucleic acids are useful for
XX identifying proteins or screening for homologous nucleic acids required
XX for cellular proliferation to isolate candidate molecules for rational
XX drug discovery programs, or for screening homologous nucleic acids
XX required for proliferation in cells other than S. aureus, S. typhimurium,
XX K. pneumoniae or P. aeruginosa. The present sequence is encoded by one of
XX the target prokaryotic essential genes. Note: The sequence data for this
XX patent did not form part of the printed specification, but was obtained
XX in electronic format directly from WIPO at
XX ftp.wipo.int/pub/published_pat_sequences
XX
XX Sequence 774 AA;
XX
XX Query Match 3.3%; Score 8; DB 6; Length 774;
XX Best Local Similarity 100.0%; Pred. No. 2.3e+02;
XX Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
```

---

```
QY 37 SAPVSCAG 44
Db 695 SAPVSCAG 702

RESULT 39
AAE32528
ID AAE32528 standard; peptide; 10 AA.
XX
XX AAE32528;
XX
XX 24-MAR-2003 (first entry)
XX
XX West nile virus (WNV) peptide #24.
XX
XX West nile virus; WNV; flavivirus-related disorder; vaccine; virucide;
XX KW immunopathogenesis; pharmacogenomic; infection; immune response.
XX
XX West nile virus.
XX
XX WO200283903-A2.
XX
XX 24-OCT-2002.
XX
XX 28-FEB-2002; 2002WO-US006575.
XX
XX 28-FEB-2001; 2001US-0272132P.
XX
XX (UYBR-) UNIV BROWN RES FOUND.
XX
XX Degroot AS, Martin W;
XX
XX WPI; 2003-103373/09.
XX
XX New West Nile Virus (WNV) vaccine, useful for inducing an anti-WNV immune
XX PT response to a mammalian subject, for treating or preventing WNV infection
XX PT or flavivirus-related disorders.
XX
XX Claim 1; Page 8; 103pp; English.
XX
XX The invention relates to vaccines comprising West nile virus (WNV)
XX vaccine candidate peptides. It also relates to methods for determining
XX WNV epitopes and methods of using the same. The vaccine is useful for
XX inducing an anti-WNV immune response to a mammalian subject, for treating
XX or preventing WNV infection or flavivirus-related disorders. The
XX candidate peptides are useful for screening exposed individuals,
XX investigating the immunopathogenesis of WNV disease in humans, as
XX components of diagnostic kits developed for the surveillance effort, or
XX as tool for measuring WNV vaccine-related immune responses. The
XX recombinant peptides and polynucleotides are useful in diagnostic assays,
XX prognostic assays, pharmacogenomics, or for monitoring clinical trials.
XX The present sequence is West nile virus peptide used in the invention
XX
XX Sequence 10 AA;
XX
XX Query Match 2.9%; Score 7; DB 6; Length 10;
XX Best Local Similarity 100.0%; Pred. No. 44;
XX Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX 21 PGLRCLN 27
XX 2 PGLRCLN 8
XX
XX RESULT 40
ABB41772
ID ABB41772 standard; peptide; 43 AA.
XX
XX ABB41772;
XX
XX 04-FEB-2002 (first entry)
XX
```

```

DE Peptide #9278 encoded by human foetal liver single exon probe.
XX Human; foetal liver; gene expression; single exon nucleic acid probe.
KW Homo sapiens.
OS WO200157277-A2.
XX 09-AUG-2001.
XX 30-JAN-2001; 2001WO-US000669.
XX 04-FEB-2000; 2000US-0180312P.
PR 26-MAY-2000; 2000US-0207456P.
PR 30-JUN-2000; 2000US-00608408.
PR 03-AUG-2000; 2000US-00632366.
PR 21-SEP-2000; 2000US-0234687P.
PR 27-SEP-2000; 2000US-0236359P.
PR 04-OCT-2000; 2000US-00024263.
XX (MOLE-) MOLECULAR DYNAMICS INC.
XX Penn SG, Hanzel DK, Chen W, Rank DR;
PI WPI; 2001-483447/52.
XX Human genome-derived single exon nucleic acid probes useful for analyzing
XX gene expression in human fetal liver.
PS Claim 27; SEQ ID NO 34407; 639pp + Sequence Listing; English.
XX The invention relates to a single exon nucleic acid probe for measuring
XX human gene expression in a sample derived from human foetal liver. The
XX single exon nucleic acid probes may be used for predicting, measuring and
XX displaying gene expression in samples derived from human fetal liver. The
XX present sequence is a peptide encoded by a single exon nucleic acid probe
XX of the invention. Note: The sequence data for this patent did not form
XX part of the printed specification, but was obtained in electronic format
XX directly from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 43 AA;
Query Match 2.9%; Score 7; DB 4; Length 43;
Best Local Similarity 100.0%; Pred.No.1.5e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 155 TMWVII 161
Db 26 TMWVII 32

RESULT 41
AAM35571
ID AAM35571 standard; protein; 43 AA.
AC AAM35571;
XX 17-OCT-2001 (first entry)
XX Peptide #9608 encoded by probe for measuring placental gene expression.
DE Probe; microarray; human; placenta; antenatal diagnosis;
XX genetic disorder.
KW Homo sapiens.
XX WO200157272-A2.
XX 09-AUG-2001.
XX 30-JAN-2001; 2001WO-US000663.
XX 04-FEB-2000; 2000US-0180312P.
PR

DE Peptide #9278 encoded by human foetal liver single exon probe.
XX Human; foetal liver; gene expression; single exon nucleic acid probe.
KW Homo sapiens.
OS WO200157277-A2.
XX 09-AUG-2001.
XX 30-JAN-2001; 2001WO-US000669.
XX 04-FEB-2000; 2000US-0180312P.
PR 26-MAY-2000; 2000US-0207456P.
PR 30-JUN-2000; 2000US-00608408.
PR 03-AUG-2000; 2000US-00632366.
PR 21-SEP-2000; 2000US-0234687P.
PR 27-SEP-2000; 2000US-0236359P.
PR 04-OCT-2000; 2000US-00024263.
XX (MOLE-) MOLECULAR DYNAMICS INC.
XX Penn SG, Hanzel DK, Chen W, Rank DR;
PI WPI; 2001-488997/53.
XX Human genome-derived single exon nucleic acid probes useful for analyzing
XX gene expression in human placenta.
PS Claim 27; SEQ ID NO 35840; 654pp; English.
XX The present invention relates to single exon nucleic acid probes (SENP;
XX see AAI31315-AAI57546). The present sequence is a peptide encoded by one
XX such probe. The probes are useful for producing a microarray for
XX predicting, measuring and displaying gene expression in samples derived
XX from human placenta. The probes are useful for antenatal diagnosis of
XX human genetic disorders
XX
SQ Sequence 43 AA;
Query Match 2.9%; Score 7; DB 4; Length 43;
Best Local Similarity 100.0%; Pred.No.1.5e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 155 TMWVII 161
Db 26 TMWVII 32

RESULT 42
AAM75460
ID AAM75460 standard; protein; 43 AA.
XX AAM75460;
XX 06-NOV-2001 (first entry)
XX Human bone marrow expressed probe encoded protein SEQ ID NO: 35766.
DE Human; bone marrow expressed exon; gene expression analysis; probe;
XX microarray; cancer; leukaemia; lymphoma; myeloma.
XX Homo sapiens.
XX WO200157276-A2.
XX 09-AUG-2001.
XX 30-JAN-2001; 2001WO-US000668.
XX 04-FEB-2000; 2000US-0180312P.
PR 26-MAY-2000; 2000US-0207456P.
PR 30-JUN-2000; 2000US-00608408.
PR 03-AUG-2000; 2000US-00632366.
PR 21-SEP-2000; 2000US-0234687P.
PR 27-SEP-2000; 2000US-0236359P.
PR 04-OCT-2000; 2000US-00024263.
XX (MOLE-) MOLECULAR DYNAMICS INC.
XX Penn SG, Hanzel DK, Chen W, Rank DR;
PI WPI; 2001-488900/53.
XX Human genome-derived single exon nucleic acid probes useful for analyzing
XX gene expression in human bone marrow.
PT

```



XX Example 4; SEQ ID NO 35766; 658pp + Sequence Listing; English.  
PS The present invention provides a number of single exon nucleic acid  
CC probes which are derived from genomic sequences expressed in the human  
CC bone marrow. They can be used to measure gene expression in bone marrow  
CC samples, which may enable the improved diagnosis and treatment of cancers  
CC such as lymphoma, leukaemia and myeloma. The present sequence is a  
CC protein encoded by one of the probes of the invention  
XX  
SQ Sequence 43 AA;  
  
Query Match 2.9%; Score 7; DB 4; Length 43;  
Best Local Similarity 100.0%; Pred. No. 1.5e+02; Mismatches 0; Indels 0; Gaps 0;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 155 TMMVIII 161  
Db 26 TMMVIII 32  
  
RESULT 43  
AAM62644  
ID AAM62644 standard; protein; 43 AA.  
AC AAM62644;  
XX  
XX 05-NOV-2001 (first entry)  
DT Human brain expressed single exon probe encoded protein SEQ ID NO: 34749.  
DE Human; brain expressed exon; gene expression analysis; probe; microarray;  
KW Alzheimer's disease; multiple sclerosis; schizophrenia; epilepsy; cancer.  
KW  
XX Homo sapiens.  
OS  
XX WO200157275-A2.  
XX  
XX 09-AUG-2001.  
XX  
XX 30-JAN-2001; 2001WO-US000667.  
XX  
XX 04-FEB-2000; 2000US-0180312P.  
PR 26-MAY-2000; 2000US-0207456P.  
PR 30-JUN-2000; 2000US-00608408.  
PR 03-AUG-2000; 2000US-00632366.  
PR 21-SEP-2000; 2000US-0234687P.  
PR 27-SEP-2000; 2000US-0236359P.  
PR 04-OCT-2000; 2000GB-00024263.  
XX  
XX (MOLE-) MOLECULAR DYNAMICS INC.  
XX  
XX Penn SG, Hanzel DK, Chen W, Rank DR;  
XX  
XX WPI; 2001-483446/52.  
XX  
XX Single exon nucleic acid probes for analyzing gene expression in human  
XX brains.  
XX  
XX Example 4; SEQ ID NO 34749; 650pp + Sequence Listing; English.  
PS The present invention provides a number of single exon nucleic acid  
CC probes which are derived from genomic sequences expressed in the human  
CC brain. They can be used to measure gene expression in brain cell samples,  
CC which may enable the diagnosis and improved treatment of nervous system  
CC diseases such as Alzheimer's disease, multiple sclerosis, schizophrenia,  
CC epilepsy and cancers. The present sequence is a protein encoded by one of  
CC the probes of the invention  
XX  
SQ Sequence 43 AA;  
  
Query Match 2.9%; Score 7; DB 4; Length 43;  
Best Local Similarity 100.0%; Pred. No. 1.5e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 155 TMMVIII 161  
Db 26 TMMVIII 32  
  
RESULT 44  
ABG57210  
ID ABG57210 standard; peptide; 43 AA.  
XX  
XX ABG57210;  
XX  
XX 25-FEB-2003 (first entry)  
DT Human liver peptide, SEQ ID No 35858.  
DE Human; liver; cirrhosis; hyperlipoproteinaemia; hyperlipidaemia;  
KW hypercholesterolaemia; coronary heart disease.  
KW  
XX Homo sapiens.  
OS  
XX WO200157273-A2.  
XX  
XX 09-AUG-2001.  
XX  
XX 30-JAN-2001; 2001WO-US000664.  
XX  
XX 04-FEB-2000; 2000US-0180312P.  
PR 26-MAY-2000; 2000US-0207456P.  
PR 30-JUN-2000; 2000US-00608408.  
PR 03-AUG-2000; 2000US-00632366.  
PR 21-SEP-2000; 2000US-0234687P.  
PR 27-SEP-2000; 2000US-0236359P.  
PR 04-OCT-2000; 2000GB-00024263.  
XX  
XX (MOLE-) MOLECULAR DYNAMICS INC.  
XX  
XX Penn SG, Hanzel DK, Chen W, Rank DR;  
XX  
XX WPI; 2001-488898/53.  
XX  
XX Human genome-derived single exon nucleic acid probes useful for analyzing  
XX gene expression in human adult liver.  
XX  
XX Claim 27; SEQ ID NO 35858; 658pp; English.  
XX  
XX The invention relates to a single exon nucleic acid probe (SENP) (I) for  
XX measuring human gene expression in a sample derived from human adult  
XX liver comprising one of 13109 defined nucleotide sequences given in the  
XX specification (or complements/ fragments). The probe hybridises at high  
XX stringency to a nucleic acid molecule expressed in the human adult liver.  
XX (I) may be used for predicting, measuring and displaying gene expression  
XX in samples derived from human adult liver. The genes identified may be  
XX involved in genetic liver diseases such as cirrhosis,  
XX hyperlipoproteinaemia, hyperlipidaemia and hypercholesterolaemia which is  
XX associated with coronary heart disease. ABG57210 represent human  
XX liver single exon encoded peptides of the invention. Note: The sequence  
XX information for this patent does not appear in the printed specification  
XX but was obtained in electronic format directly from WIPO at  
XX ftp.wipo.int/pub/published\_pct\_sequences  
XX  
SQ Sequence 43 AA;  
  
Query Match 2.9%; Score 7; DB 4; Length 43;  
Best Local Similarity 100.0%; Pred. No. 1.5e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

RESULT 45  
ABG45027  
ID ABG45027 standard; peptide; 43 AA.  
XX  
AC ABG45027;  
XX  
DT 19-AUG-2002 (first entry)  
XX  
DE Human peptide encoded by genome-derived single exon probe SEQ ID 34592.  
XX  
KW Human; single exon probe; asthma; lung cancer; COPD; ILD;  
KW chronic obstructive pulmonary disease; interstitial lung disease;  
KW familial idiopathic pulmonary fibrosis; neurofibromatosis;  
KW tuberous sclerosis; Gaucher's disease; Niemann-Pick disease;  
KW Hermansky-Pudlak syndrome; sarcoidosis; pulmonary haemorrhoidosis;  
KW pulmonary histiocytosis; lymphangioleiomyomatosis; Karagener syndrome;  
KW pulmonary alveolar proteinosis; fibrocystic pulmonary dysplasia;  
KW primary ciliary dyskinesia; pulmonary hypertension;  
KW hyaline membrane disease.  
XX  
OS Homo sapiens.  
XX  
XX WO200186003-A2.  
XX  
XX 15-NOV-2001.  
XX  
XX 30-JAN-2001; 2001WO-US000665.  
XX  
XX 04-FEB-2000; 2000US-0180312P.  
XX  
XX 26-MAY-2000; 2000US-0207456P.  
XX  
XX 30-JUN-2000; 2000US-00608408.  
XX  
XX 03-AUG-2000; 2000US-00832366.  
XX  
XX 21-SEP-2000; 2000US-0234687P.  
XX  
XX 27-SEP-2000; 2000US-0236359P.  
XX  
XX 04-OCT-2000; 2000GB-00024263.  
XX  
XX (MOLE-) MOLECULAR DYNAMICS INC.  
XX  
XX Penn SG, Hanzel DK, Chen W, Rank DR;  
XX  
XX WPI; 2002-114183/15.  
XX  
XX Spatially-addressable set of single exon nucleic acid probes, used to  
XX measure gene expression in human lung samples.  
XX  
XX Claim 27; SEQ ID NO 34692; 634pp; English.  
XX  
XX The invention relates to a spatially-addressable set of single exon  
XX nucleic acid probes for measuring gene expression in a sample derived  
XX from human lung comprising single exon nucleic acid probes having one of  
XX 12614 nucleic acid sequences mentioned in the specification, or their  
XX complements or the 12387 open reading frames derived from the 12614  
XX probes. Also included are a microarray comprising the novel set of probes  
XX; the novel set of probes which hybridise at high stringency to a nucleic  
XX acid expressed in the human lung; measuring gene expression in a sample  
XX derived from human lung, comprising (a) contacting the array with a  
XX collection of detectably labeled nucleic acids derived from human lung  
XX mRNA, and (b) measuring the label detectably bound to each probe of the  
XX array; identifying exons in a eukaryotic genome, comprising (a)  
XX algorithmically predicting at least one exon from genomic sequences of  
XX the eukaryote; and (b) detecting specific hybridisation of detectably  
XX labeled nucleic acids from eukaryote lung mRNA, to a single exon probe,  
XX having a fragment identical to the predicted exon, the probe is included  
XX in the above mentioned microarray; assigning exons to a single gene,  
XX comprising (a) identifying exons from genomic sequence by the method  
XX above and (b) measuring the expression of each of the exons in several  
XX tissues and/or cell types using hybridisation to a single exon  
XX microarrays having a probe with the exon, where a common pattern of  
XX expression of the exons in the tissues and/or cell types indicates that  
XX the exons should be assigned to a single gene; a peptide comprising one  
XX of 12011 sequences, mentioned in the specification, or encoded by the  
XX probes/open reading frames (ORF). The probes are used for gene expression  
XX analysis, and for identifying exons in a gene, particularly using human

CC lung derived mRNA and for the study of lung diseases such as asthma, lung  
CC cancer, chronic obstructive pulmonary disease (COPD), interstitial lung  
CC disease (ILD), familial idiopathic pulmonary fibrosis, neurofibromatosis,  
CC tuberous sclerosis, Gaucher's disease, Niemann-Pick disease, Hermansky-  
CC Pudlak syndrome, sarcoidosis, pulmonary haemorrhoidosis, pulmonary  
CC histiocytosis, lymphangioleiomyomatosis, pulmonary alveolar proteinosis,  
CC Karagener syndrome, fibrocystic pulmonary dysplasia, primary ciliary  
CC dyskinesia, pulmonary hypertension and hyaline membrane disease. The  
CC present sequence is a peptide/protein encoded by a single exon probe of  
CC the invention. Note: The sequence data for this patent did not form part  
CC of the printed specification, but was obtained in electronic format  
CC directly from WIPO at ftp.wipo.int/pub/published\_pct\_sequences  
XX  
XX Sequence 43 AA;  
SQ  
Query Match 2.9%; Score 7; DB 5; Length 43;  
Best Local Similarity 100.0%; Pred. No. 1.5e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 155 TMWVWIII 161  
DB 26 TMWVWIII 32  
|||||  
|||||  
RESULT 46  
ABG99077  
ID ABB99077 standard; protein; 83 AA.  
XX  
AC ABB99077;  
XX  
XX 07-FEB-2003 (first entry)  
XX  
XX AP-2gamma transcription factor 9.13 amino acid sequence.  
XX  
XX AP-2gamma transcription factor 9.13; embryo development disorder.  
XX  
XX Unidentified.  
XX  
XX CN1351020-A.  
XX  
XX 29-MAY-2002.  
XX  
XX 26-OCT-2000; 2000CN-00125786.  
XX  
XX 26-OCT-2000; 2000CN-00125786.  
XX  
XX (BODE-) BODE GENE DEV CO LTD SHANGHAI.  
XX  
XX PI Mao Y, Xie Y;  
XX  
XX WPI; 2002-644401/70.  
XX  
XX N-PSDB; ABV75472.  
XX  
XX Polypeptide-AP-2 gamma transcription factor 9.13.  
XX  
XX Claim 1; Page 26-27 (disclosure); 33pp; Chinese.  
XX  
XX The invention relates to a new polypeptide designated AP-2 gamma  
XX transcription factor 9.13. The application of the polypeptide is in  
XX treating diseases such as embryo development disorder. The antagonist of  
XX the polypeptide and its medical action are also disclosed. The current  
XX sequence represents the AP-2gamma transcription factor 9.13 amino acid  
XX sequence  
XX  
XX Sequence 83 AA;  
SQ  
Query Match 2.9%; Score 7; DB 5; Length 83;  
Best Local Similarity 100.0%; Pred. No. 2.7e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 143 KDLGTGLG 149  
DB 76 KDLGTGLG 82  
|||||  
|||||

## RESULT 47

ABR55213  
ID ABR55213 standard; protein; 111 AA.

AC ABR55213;

DT 29-JUL-2003 (first entry)

DE FtsZ fragment encoded by a maize EST.

XX FtsZ; starch granule; starch-producing plant; plant;

KW expressed sequence tag; EST.

OS Zea mays.

XX WO2003035874-A1.

XX 01-MAY-2003.

XX 24-OCT-2002; 2002WO-GB004806.

XX 24-OCT-2001; 2001GB-00025493.

PR 09-JAN-2002; 2002US-0346905P.

XX (GEMS-) GEMSTAR CAMBRIDGE LTD.

PI Coates SA, Burrell MM;

XX WPI; 2003-449270/42.

DR N-PSDB; ACC43934.

XX New isolated FtsZ nucleic acid or protein, useful for altering starch granules in a plant.

PS Claim 7; Page 118-119; 130pp; English.

XX The present sequence is encoded by an expressed sequence tag (EST), which is a fragment of maize FtsZ. FtsZ polypeptides alter the size and quantity of starch granules in a plant. FtsZ polynucleotides are useful for altering any one of starch granule number, starch granule size, and starch granule distribution in a starch-producing plant. By altering aspects of starch related to starch granules, the starch extracted from these transgenic plants may be more favourable for nutritional or industrial uses

XX SQ Sequence 111 AA;

Query Match 2.9%; Score 7; DB 6; Length 111;  
Best Local Similarity 100.0%; Pred. No. 3.5e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 122 ARAVQPV 128

DB 94 ARAVQPV 100

## RESULT 48

AAU46725  
ID AAU46725 standard; protein; 121 AA.

XX AAU46725;

XX 27-FEB-2002 (first entry)

XX Propionibacterium acnes immunogenic protein #7621.

XX SAPHO syndrome; synovitis; acne; pustulosis; hypertosis; osteomyelitis;  
KW uveitis; endophthalmitis; bone; joint; central nervous system; ELISA;  
XX inflammatory lesion; acne vulgaris; enzyme linked immunosorbent assay;  
XX dermatological; osteopathic; neuroprotectant.

OS Propionibacterium acnes.

XX WO200181581-A2.

XX 01-NOV-2001.

XX 20-APR-2001; 2001WO-US012865.

XX 21-APR-2000; 2000US-0199047P.

PR 02-JUN-2000; 2000US-0208841P.

PR 07-JUL-2000; 2000US-0216747P.

XX (CORI-) CORIXA CORP.

XX Skeiky YAW, Persing DH, Mitcham JL, Wang SS, Bhatia A;

PI L'maisonneuve J, Zhang Y, Jen S, Carter D;

XX WPI; 2001-616774/71.

DR N-PSDB; AAS59535.

XX Propionibacterium acnes polypeptides and nucleic acids useful for vaccinating against and diagnosing infections, especially useful for treating acne vulgaris.

XX Example 1; SEQ ID NO 7920; 1069pp; English.

XX Sequences AAU39105-AAU68017 represent Propionibacterium acnes immunogenic polypeptides. The proteins and their associated DNA sequences are used in the treatment, prevention and diagnosis of medical conditions caused by P. acnes. The disorders include SAPHO syndrome (synovitis, acne, pustulosis, hypertosis and osteomyelitis), uveitis and endophthalmitis. P. acnes is also involved in infections of bone, joints and the central nervous system, however it is particularly involved in the inflammatory lesions associated with acne vulgaris. A method for detecting the presence or absence of P. acnes in a patient comprises contacting a sample with a binding agent that binds to the proteins of the invention and determining the amount of bound protein in the sample. The polypeptides may be used as antigens in the production of antibodies specific for P. acnes proteins. These antibodies can be used to downregulate expression and activity of P. acnes polypeptides and therefore treat P. acnes infections. The antibodies may also be used as diagnostic agents for determining P. acnes presence, for example, by enzyme linked immunosorbent assay (ELISA). Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published\_pct\_sequences

XX SQ Sequence 121 AA;

Query Match 2.9%; Score 7; DB 4; Length 121;  
Best Local Similarity 100.0%; Pred. No. 3.7e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 84 TSQALPA 90

DB 1 TSQALPA 7

## RESULT 49

ABM43244

ID ABM43244 standard; protein; 121 AA.

XX ABM43244;

XX 20-OCT-2003 (first entry)

XX Propionibacterium acnes predicted ORF-encoded polypeptide #7920.

XX Acne vulgaris; antisephorhoic; dermatological; antibacterial;  
KW immunostimulant; immune response; vaccine.

XX Propionibacterium acnes.

FN WO2003033515-A1.  
 PD 24-APR-2003.  
 XX 11-OCT-2002; 2002WO-US032727.  
 XX 15-OCT-2001; 2001US-00978825.  
 PR (CORI-) CORIXA CORP.  
 PA Mitcham JL, Skeiky YAW, Persing DH, Bhatia A, Maisonneuve JL;  
 PI Zhang Y, Wang S, Jen S, Lodes MJ, Benson DR, Jones R, Carter D;  
 PI Barth B, Vallie-Douglass J;  
 XX WPI; 2003-381789/36.  
 DR N-PSDB; ACF64464.  
 XX New Protonibacterium acnes polypeptides and polynucleotides encoding the  
 PT polypeptide, useful for diagnosing, preventing or treating acne vulgaris,  
 PT or for stimulating an immune response specific for a P. acnes protein.  
 XX  
 PS Example 1; SEQ ID NO 7920; 1481pp; English.  
 XX  
 CC The invention relates to an isolated polynucleotide (ACF64435-ACF64733)  
 CC encoding a Protonibacterium acnes protein. The invention also relates to  
 CC polypeptides encoded by the polynucleotides (ABM35624-ABM64536) and to  
 CC immunogenic fragments of P. acnes polypeptides. The invention  
 CC additionally encompasses expression vectors and host cells comprising a  
 CC polynucleotide of the invention; antibodies against polypeptides of the  
 CC polynucleotide of the invention; a polypeptide of the invention; a  
 CC method for stimulating an immune response specific for a P. acnes  
 CC polypeptide and an isolated T cell population comprising P. acnes polypeptides,  
 CC via this method; a vaccine composition (comprising P. acnes polypeptides,  
 CC polynucleotides, antibodies, fusion proteins, T cell populations, or  
 CC antigen-presenting cells that express the polypeptide); a method and kit  
 CC for detecting or determining the presence or absence of P. acnes in a  
 CC patient; and a method for inhibiting the development of P. acnes in a  
 CC patient. The P. acnes polypeptides, polynucleotides, antibodies, fusion  
 CC proteins, T cell populations or antigen-presenting cells that express the  
 CC polypeptides are useful for diagnosing, preventing or treating acne  
 CC vulgaris, or for stimulating an immune response specific for a P. acnes  
 CC protein. The polynucleotides can also be used as probes or primers for  
 CC nucleic acid hybridisation. The vaccine composition is useful for the  
 CC stimulation of an immune response against P. acnes, or for treating acne,  
 CC and the kit is useful for performing a diagnostic assay. The present  
 CC sequence represents a polypeptide predicted to be encoded by an ORF (open  
 CC reading frame) contained within the P. acnes polynucleotides of the  
 CC invention. Note: The sequence data for this patent did not form part of  
 CC the printed specification, but was obtained in electronic format directly  
 CC from WIPO at ftp.wipo.int/pub/published\_pct\_sequences  
 XX  
 SQ Sequence 121 AA;  
 Query Match 2.9%; Score 7; DB 6; Length 121;  
 Best Local Similarity 100.0%; Pred. No. 3.7e+02;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 84 TSQALPA 90  
 DB 1 TSQALPA 7  
 RESULT 50  
 ID ABP07491  
 AC ABP07491 standard; protein; 133 AA.  
 AC ABP07491;  
 XX 25-JUN-2002 (first entry)  
 DT Human ORFX protein sequence SEQ ID NO:14964.  
 XX Human; open reading frame; ORFX; gene therapy; cancer; cirrhosis;  
 KW

KW hyperproliferative disorder; psoriasis; benign tumour; haemorrhage;  
 KW degenerative disorder; osteoarthritis; neurodegenerative disorder;  
 KW cardiovascular disease; diabetes mellitus; systemic lupus erythematosus;  
 KW hypertension; hypothyroidism; cholesterol ester storage disease;  
 KW immune deficiency; immune disorder; infectious disease;  
 KW autoimmune disorder; rheumatoid arthritis; autoimmune thyroiditis;  
 KW myasthenia gravis.  
 XX Homo sapiens.  
 XX WO200192523-A2.  
 XX 06-DEC-2001.  
 XX 29-MAY-2001; 2001WO-US010836.  
 XX 30-MAY-2000; 2000US-0206132P.  
 PR 29-AUG-2000; 2000US-0228716P.  
 XX (CURA-) CURAGEN CORP.  
 XX Shimketa RA, Leach MD;  
 PI WPI; 2002-106308/14.  
 DR N-PSDB; ABN23243.  
 XX Novel human polypeptides and polynucleotides useful for diagnosing,  
 PT preventing and treating cardiovascular disease, neurodegenerative,  
 PT hyperproliferative disorders and autoimmune disorders.  
 XX  
 PS Disclosure; SEQ ID NO 14964; 1037pp; English.  
 XX  
 CC The present invention describes substantially purified human proteins  
 CC (referred to as open reading frame, ORFX, where X is 1-11491 (see Table 1  
 CC in the specification). ABN15762 to ABN27252 encode the human ORFX  
 CC proteins given in ABP0010 to ABP11500. ORFX proteins are useful for  
 CC treating or preventing a pathology associated with an ORFX-associated  
 CC disorder in humans, and in the manufacture of a medicament for treating a  
 CC syndrome associated with ORFX-associated disorder. ORFX polynucleotide  
 CC sequences can be used in gene therapy. ORFX sequences can be used in the  
 CC treatment of cancer, hyperproliferative disorders, cirrhosis of liver,  
 CC psoriasis, benign tumours, keloid, degenerative disorders, haemorrhage,  
 CC osteoarthritis, neurodegenerative disorders, diabetes mellitus, systemic  
 CC transplantation, cardiovascular diseases, diabetes mellitus, systemic  
 CC lupus erythematosus, hypertension, hypothyroidism, cholesterol ester  
 CC storage disease, various immune deficiencies and disorders, infectious  
 CC diseases, autoimmune disorders such as multiple sclerosis, rheumatoid  
 CC arthritis, autoimmune thyroiditis, myasthenia gravis, graft-versus-host  
 CC disease and autoimmune inflammatory eye disease. ORFX proteins are also  
 CC useful for treating burns, incisions, ulcers, for treating osteoporosis,  
 CC bone degenerative disorders, or periodontal disease, and for gut  
 CC protection or regeneration and treatment of lung or liver fibrosis,  
 CC reperfusion injury in various tissues and conditions resulting from  
 CC systemic cytokine damage. N.B. The sequence data for this patent did not  
 CC form part of the printed specification, but was obtained in electronic  
 CC format directly from WIPO at ftp.wipo.int/pub/published\_pct\_sequences  
 XX  
 SQ Sequence 133 AA;  
 Query Match 2.9%; Score 7; DB 5; Length 133;  
 Best Local Similarity 100.0%; Pred. No. 4e+02;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 100 EGFGADE 106  
 DB 101 EGFGADE 107  
 RESULT 51  
 ID ABG04700  
 AC ABG04700 standard; protein; 140 AA.  
 XX  
 AC ABG04700;

XX 13-FEB-2002 (first entry)  
 XX Novel human diagnostic protein #4691.  
 DE Human; chromosome mapping; gene mapping; gene therapy; forensic;  
 XX food supplement; medical imaging; diagnostic; genetic disorder.  
 KW Homo sapiens.  
 XX WO200175067-A2.  
 XX 11-OCT-2001.  
 PD 30-MAR-2001; 2001WO-US008631.  
 XX 31-MAR-2000; 2000US-00540217.  
 PR 23-AUG-2000; 2000US-00649167.  
 XX (HYSE-) HYSEQ INC.  
 XX Drmanac RT, Liu C, Tang YT;  
 XX WPI; 2001-639362/73.  
 DR N-PSDB; AAS68887.  
 XX New isolated polynucleotide and encoded polypeptides, useful in  
 PT diagnostics, forensics, gene mapping, identification of mutations  
 PT responsible for genetic disorders or other traits and to assess  
 PT biodiversity.  
 XX Claim 20; SEQ ID NO 35059; 103pp; English.  
 PS The invention relates to isolated polynucleotide (I) and polypeptide (II)  
 CC sequences. (I) is useful as hybridisation probes, polymerase chain  
 CC reaction (PCR) primers, oligomers, and for chromosome and gene mapping,  
 CC and in recombinant production of (II). The polynucleotides are also used  
 CC in diagnostics as expressed sequence tags for identifying expressed  
 CC genes. (I) is useful in gene therapy techniques to restore normal  
 CC activity of (II) or to treat disease states involving (II). (II) is  
 CC useful for generating antibodies against it, detecting or quantitating a  
 CC polypeptide in tissue, as molecular weight markers and as a food  
 CC supplement. (II) and its binding partners are useful in medical imaging  
 CC of sites expressing (II). (I) and (II) are useful for treating disorders  
 CC involving aberrant protein expression or biological activity. The  
 CC polypeptide and polynucleotide sequences have applications in  
 CC diagnostics, forensics, gene mapping, identification of mutations  
 CC responsible for genetic disorders or other traits to assess biodiversity  
 CC and to produce other types of data and products dependent on DNA and  
 CC amino acid sequences. ABG00010-ABG30377 represent novel human diagnostic  
 CC amino acid sequences of the invention. Note: The sequence data for this  
 CC patent did not appear in the printed specification, but was obtained in  
 CC electronic format directly from WIPO at  
 CC ftp.wipo.int/pub/published\_pct\_sequences  
 XX Sequence 140 AA;  
 SQ Query Match 2.9%; Score 7; DB 4; Length 140;  
 Best Local Similarity 100.0%; Pred. No. 4.2e+02;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 98 ASEGPGA 104  
 Db 10 ASEGPGA 16  
 RESULT 52  
 ABU31178  
 ID ABU31178 standard; protein; 153 AA.  
 XX AC ABU31178;  
 XX 19-JUN-2003 (first entry)  
 DT

XX Protein encoded by Prokaryotic essential gene #16705.  
 DE Antisense; prokaryotic essential gene; cell proliferation; drug design.  
 KW Klebsiella pneumoniae.  
 XX WO200277183-A2.  
 XX 03-OCT-2002.  
 PD 21-MAR-2002; 2002WO-US009107.  
 XX 21-MAR-2001; 2001US-00815242.  
 PR 06-SEP-2001; 2001US-00948993.  
 XX 25-OCT-2001; 2001US-0342523P.  
 PR 08-FEB-2002; 2002US-00072851.  
 PR 06-MAR-2002; 2002US-0362699P.  
 XX (ELIT-) ELITRA PHARM INC.  
 XX Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;  
 PI Wall D, Trawick JD, Carr GU, Yamamoto R, Forsyth RA, Xu HH;  
 XX WPI; 2003-029926/02.  
 DR N-PSDB; ACA35048.  
 XX New antisense nucleic acids, useful for identifying proteins or screening  
 PT for homologous nucleic acids, required for cellular proliferation to  
 PT isolate candidate molecules for rational drug discovery programs.  
 XX Claim 25; SEQ ID NO 59102; 1766pp; English.  
 PS The invention relates to an isolated nucleic acid comprising any one of  
 CC the 6213 antisense sequences given in the specification where expression  
 CC of the nucleic acid inhibits proliferation of a cell. Also included are:  
 CC (1) a vector comprising a promoter operably linked to the nucleic acid  
 CC encoding a polypeptide whose expression is inhibited by the antisense  
 CC nucleic acid; (2) a host cell containing the vector; (3) an isolated  
 CC polypeptide or its fragment whose expression is inhibited by the  
 CC antisense nucleic acid; (4) an antibody capable of specifically binding  
 CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular  
 CC proliferation or the activity of a gene in an operon required for  
 CC proliferation; (7) identifying a compound that influences the activity of  
 CC the gene product or that has an activity against a biological pathway  
 CC required for proliferation, or that inhibits cellular proliferation; (8)  
 CC identifying a gene required for cellular proliferation or the biological  
 CC pathway in which a proliferation-required gene or its gene product lies  
 CC or a gene on which the test compound that inhibits proliferation of an  
 CC organism acts; (9) manufacturing an antibiotic; (10) profiling a  
 CC compound's activity; (11) a culture comprising strains in which the gene  
 CC product is overexpressed or underexpressed; (12) determining the extent  
 CC to which each of the strains is present in a culture or collection of  
 CC strains; or (13) identifying the target of a compound that inhibits the  
 CC proliferation of an organism. The antisense nucleic acids are useful for  
 CC identifying proteins or screening for homologous nucleic acids required  
 CC for cellular proliferation to isolate candidate molecules for rational  
 CC drug discovery programs, or for screening homologous nucleic acids  
 CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,  
 CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is encoded by one of  
 CC the target prokaryotic essential genes. Note: The sequence data for this  
 CC patent did not form part of the printed specification, but was obtained  
 CC in electronic format directly from WIPO at  
 CC ftp.wipo.int/pub/published\_pct\_sequences  
 XX Sequence 153 AA;  
 SQ Query Match 2.9%; Score 7; DB 6; Length 153;  
 Best Local Similarity 100.0%; Pred. No. 4.6e+02;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 18 SPACJR 24  
 |||||

```

Db      16 SPAGLR 22

RESULT 53
AAB70881
ID AAB70881 standard; protein; 154 AA.
XX
AC AAB70881;
XX
DT 12-JUL-2001 (first entry)
XX
DE C. glutamicum lrp protein.
XX
KW Fermentation; L-amino acid production; lrp gene; lysine; isoleucine;
KW medicine; animal feed supplement.
XX
OS Corynebacterium glutamicum.
XX
PN EP1090993-A1.
XX
PD 11-APR-2001.
XX
PF 29-SEP-2000; 2000EP-00121159.
XX
PR 05-OCT-1999; 99DE-01047792.
XX
PA (DEGS ) DEGUSSA-HUELS AG.
XX
PI Moeckel B, Pfefferle W, Puhler A, Kalinowski J, Bathe B;
XX
DR WPI; 2001-292927/31.
DR N-PSDB; AAF61688.
XX
PT New lrp gene from coryneform bacteria, used to prepare transformants with
PT increased synthesis of amino acids, particularly lysine and isoleucine.
XX
PS Claim 6; Page 15; 22pp; German.
XX
CC This invention describes a novel isolated nucleic acid (I) from
CC coryneform bacteria used for the fermentative production of selected L-
CC amino acids, by fermenting the amino acid-producing coryneform in which
CC at least the lrp gene has been weakened or amplified, then isolating
CC amino acids that have accumulated in the medium or cells. (I) is used to
CC transform coryneforms for production of L-amino acids, specifically
CC lysine and isoleucine, which are used in medicine and particularly as
CC animal feed supplement. It may also be used as probes and primers for
CC isolating related sequences. Regulating expression of (I) improves
CC production of amino acids, especially of L-lysine. This sequence
CC represents the Corynebacterium glutamicum lrp protein which is used in
CC the method described in the invention
XX
SQ Sequence 154 AA;

Query Match 2.9%; Score 7; DB 4; Length 154;
Best Local Similarity 100.0%; Pred. No. 4.6e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 166 GIILGYS 172
Db 50 GIILGYS 56

RESULT 54
AAG90034
ID AAG90034 standard; protein; 154 AA.
XX
AC AAG90034;
XX
DT 26-SEP-2001 (first entry)
XX
DE C glutamicum protein fragment SEQ ID NO: 3788.
XX
KW Coryneform bacterium; amino acid synthesis; vitamin; saccharide;

```

```

KW organic acid synthesis.
XX
OS Corynebacterium glutamicum.
XX
PN EP1108790-A2.
XX
PD 20-JUN-2001.
XX
PF 18-DEC-2000; 2000EP-00127688.
XX
PR 16-DEC-1999; 99JP-00377484.
PR 07-APR-2000; 2000JP-00159162.
PR 03-AUG-2000; 2000JP-00280988.
XX
PA (KYOW ) KYOWA HAKKO KOGYO KK.
XX
PI Nakagawa S, Mizoguchi H, Ando S, Hayashi M, Ochiai K, Yokoi H;
PI Tateishi N, Senoh A, Ikeda M, Ozaki A;
XX
DR WPI; 2001-376931/40.
DR N-PSDB; AAH65253.
XX
PT Novel polynucleotides derived from Coryneform bacteria, for identifying
PT mutation point of a gene, measuring expression of a gene, analyzing
PT expression profile or pattern of a gene and identifying homologous gene.
XX
PS Claim 17; SEQ ID NO 3788; 246pp + Sequence Listing; English.
XX
CC The present invention provides a number of nucleotide and protein
CC sequences from the Coryneform bacterium Corynebacterium glutamicum. These
CC are useful for identifying the mutation point of a gene derived from a
CC mutant of coryneform bacterium, measuring expression amount and analysing
CC the expression profile or expression pattern of a gene derived from
CC Coryneform bacterium, and identifying a homologue of a gene derived from
CC Coryneform bacterium. Coryneform bacteria are useful for producing amino
CC acids, nucleic acids, vitamins, saccharides and organic acids,
CC particularly L-lysine. The present sequence is a protein described in the
CC exemplification of the invention. Note: The sequence data for this patent
CC did not form part of the printed specification, but was obtained in
CC electronic format directly from the European Patent Office
XX
SQ Sequence 154 AA;

Query Match 2.9%; Score 7; DB 4; Length 154;
Best Local Similarity 100.0%; Pred. No. 4.6e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 166 GIILGYS 172
Db 50 GIILGYS 56

RESULT 55
AAU32517
ID AAU32517 standard; protein; 181 AA.
XX
AC AAU32517;
XX
DT 18-DEC-2001 (first entry)
XX
DE Novel human secreted protein #3008.
XX
KW Human; vaccination; gene therapy; nutritional supplement;
KW stem cell proliferation; haematopoiesis; nerve tissue regeneration;
KW immune suppression; immune stimulation; anti-inflammatory; leukaemia.
XX
OS Homo sapiens.
XX
PN WO200179449-A2.
XX
PD 25-OCT-2001.
XX
PF 16-APR-2001; 2001WO-US008656.

```

XX 18-APR-2000; 2000US-00552929.  
 PR 26-JAN-2001; 2001US-00770160.  
 XX (HYSB-) HYSEQ INC.  
 XX PA  
 XX PI Tang YT, Liu C, Drmanac RT;  
 XX DR WPI; 2001-611725/70.  
 XX PT Nucleic acids encoding a range of human polypeptides, useful in genetic  
 XX PT vaccination, testing and therapy.  
 XX PS Claim 20; Page 625-626; 765pp; English.  
 XX CC The invention relates to novel human secreted polypeptides. The  
 CC polypeptides and antibodies to the polypeptides are useful for  
 CC determining the presence of or predisposition to a disease associated  
 CC with altered levels of polypeptide. The polypeptides are also useful for  
 CC identifying agents (agonists and antagonists) that bind to them. Cells  
 CC expressing the proteins are useful for identifying a therapeutic agent  
 CC for use in treatment of a pathology related to aberrant expression or  
 CC physiological interactions of the polypeptide. Vectors comprising the  
 CC nucleic acids encoding the polypeptides and cells genetically engineered  
 CC to express them are also useful for producing the proteins. The proteins  
 CC are useful in genetic vaccination, testing and therapy, and can be used  
 CC as nutritional supplements. They may be used to increase stem cell  
 CC proliferation; to regulate haematopoiesis; and in bone, cartilage, tendon  
 CC and/or nerve tissue growth or regeneration; immune suppression and/or  
 CC stimulation; as anti-inflammatory agents; and in treatment of leukaemias.  
 CC AAU29510-AAU3304 represent the amino acid sequences of novel human  
 CC secreted proteins of the invention  
 XX SQ Sequence 181 AA;

Query Match 2.9%; Score 7; DB 4; Length 181;  
 Best Local Similarity 100.0%; Pred. No. 5.3e-02;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 19 PAPGLRC 25  
 |||||  
 Db 117 PAPGLRC 123

RESULT 56  
 AAY35428  
 ID AAY35428 standard; protein; 182 AA.

XX AAY35428;  
 XX AC  
 XX DT 17-OCT-2003 (revised)  
 XX DT 13-SEP-1999 (first entry)  
 XX DE Chlamydia pneumoniae transmembrane protein sequence.  
 XX KW Respiratory disease; pneumonia; bronchitis; heart disease; sarcoidosis;  
 KW sinusitis; purulent otitis media; erythema nodosum; pharyngitis; vaccine;  
 KW neutralising epitope.  
 XX OS Chlamydoiphila pneumoniae.  
 XX FN WO9927105-A2.  
 XX PD 03-JUN-1999.  
 XX PF 20-NOV-1998; 98WO-IB001890.  
 XX PR 21-NOV-1997; 97FR-00014573.  
 XX PR 04-NOV-1998; 98US-0107078P.  
 XX PA (GBST ) GENSET.  
 XX PI Griffiths R;

XX WPI; 1999-357842/30.  
 XX DR Genome sequence of Chlamydia pneumoniae.  
 XX PT  
 XX PS Page 1210; Disclosure; 1912pp; English.  
 XX CC AAY34584-Y35879 represent the proteins encoded by all the open reading  
 CC frames in the complete genome (see AAY31990) of Chlamydia pneumoniae. C.  
 CC pneumoniae causes respiratory disease such as pneumonia and bronchitis  
 CC and is thought to be a contributing factor in heart disease, sarcoidosis,  
 CC sinusitis, purulent otitis media, erythema nodosum or pharyngitis. The  
 CC polypeptides encoded by the open reading frames of the C. pneumoniae  
 CC genome (see AAY34584-Y35879) can be used in immunogenic compositions as  
 CC vaccines. Vectors containing C. pneumoniae nucleotide sequences can also  
 CC be used as immunogenic compositions, especially where the vector directs  
 CC the expression of a neutralising epitope of C. pneumoniae. (Updated on 17  
 CC -OCT-2003 to standardise OS field)  
 XX SQ Sequence 182 AA;

Query Match 2.9%; Score 7; DB 2; Length 182;  
 Best Local Similarity 100.0%; Pred. No. 5.3e-02;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 143 KDLGTLG 149  
 |||||  
 Db 42 KDLGTLG 48

RESULT 57  
 ABP09730  
 ID ABP09730 standard; protein; 196 AA.

XX ABP09730;  
 XX AC  
 XX DT 25-JUN-2002 (first entry)  
 XX DE Human ORFX protein sequence SEQ ID NO:19442.  
 XX KW Human; open reading frame; ORFX; gene therapy; cancer; cirrhosis;  
 KW hyperproliferative disorder; psoriasis; benign tumour; haemorrhage;  
 KW degenerative disorder; osteoarthritis; neurodegenerative disorder;  
 KW cardiovascular disease; diabetes mellitus; systemic lupus erythematosus;  
 KW hypertension; hypothyroidism; cholesterol ester storage disease;  
 KW immune deficiency; immune disorder; infectious disease;  
 KW autoimmune disorder; rheumatoid arthritis; autoimmune thyroiditis;  
 KW myasthenia gravis.  
 XX OS Homo sapiens.  
 XX FN WO200192523-A2.  
 XX PD 06-DEC-2001.  
 XX PF 29-MAY-2001; 2001WO-US010836.  
 XX PR 30-MAY-2000; 2000US-0206132P.  
 XX PR 29-AUG-2000; 2000US-0228716P.  
 XX PA (CURA-) CURAGEN CORP.  
 XX PI Shinkets RA, Leach MD;  
 XX DR WPI; 2002-106308/14.  
 XX DR N-PSDB; ABN25482.  
 XX CC Novel human polypeptides and polynucleotides useful for diagnosing,  
 CC preventing and treating cardiovascular disease, neurodegenerative,  
 CC hyperproliferative disorders and autoimmune disorders.  
 XX PS Disclosure; SEQ ID NO 19442; 1037pp; English.

CC The present invention describes substantially purified human proteins  
 CC (referred to as open reading frame, ORFX, where X is 1-11491 (see Table 1  
 CC in the specification). ABN15762 to ABN27252 encode the human ORFX  
 CC proteins given in BP00010 to ABP11500. ORFX proteins are useful for  
 CC treating or preventing a pathology associated with an ORFX-associated  
 CC disorder in humans, and in the manufacture of a medicament for treating a  
 CC syndrome associated with ORFX-associated disorder. ORFX polynucleotide  
 CC sequences can be used in gene therapy. ORFX sequences can be used in the  
 CC treatment of cancer, hyperproliferative disorders, cirrhosis of liver,  
 CC psoriasis, benign tumors, keloid, degenerative disorders, haemorrhage,  
 CC osteoarthritis, neurodegenerative disorders, disorders related to organ  
 CC transplantation, cardiovascular diseases, diabetes mellitus, systemic  
 CC lupus erythematosus, hypertension, hypothyroidism, cholesterol ester  
 CC storage disease, various immune deficiencies and disorders, infectious  
 CC diseases, autoimmune disorders such as multiple sclerosis, rheumatoid  
 CC arthritis, autoimmune thyroiditis, myasthenia gravis, graft-versus-host  
 CC disease and autoimmune inflammatory eye disease. ORFX proteins are also  
 CC useful for treating burns, incisions, ulcers, for treating osteoporosis,  
 CC bone degenerative disorders, or periodontal disease, and for gut  
 CC protection or regeneration and treatment of lung or liver fibrosis,  
 CC reperfusion injury in various tissues and conditions resulting from  
 CC systemic cytokine damage. N.B. The sequence data for this patent did not  
 CC form part of the printed specification, but was obtained in electronic  
 CC format directly from WIPO at ftp.wipo.int/pub/published\_pct\_sequences  
 XX  
 SQ Sequence 196 AA;

Query Match 2.9%; Score 7; DB 5; Length 196;  
 Best Local Similarity 100.0%; Pred. No. 5.6e+02;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 223 PVDPOEG 229  
 |||||  
 DB 27 PVDPOEG 33

RESULT 58  
 ADC96351  
 ID ADC96351 standard; protein; 204 AA.

XX AC ADC96351;  
 XX  
 DT 01-JAN-2004 (first entry)  
 DE E. faecium protein sequence SEQ ID 5978.  
 KW Vaccine; urinary tract infection; bacteraemia; endocarditis; wound;  
 KW abdominal-pelvic infection.

XX OS Enterococcus faecium.  
 XX PN US6583275-B1.  
 XX PD 24-JUN-2003.  
 XX PF 30-JUN-1998; 98US-00107532.  
 XX PR 02-JUL-1997; 97US-0051571P.  
 XX PR 14-MAY-1998; 98US-0085598P.  
 XX PA (GENO-) GENOME THERAPEUTICS CORP.  
 XX  
 XX PI Doucette-Stamm LA, Bush D;  
 XX DR WPI; 2003-799836/75.  
 XX DR N-PSDB; ADC92697.

XX PT New isolated nucleic acid derived from Enterococcus faecium encoding an  
 XX PT Enterococcus faecium polypeptide useful for detection, prevention and  
 XX PT treatment of a pathological condition resulting from a bacterial  
 XX PT infection.  
 XX  
 XX Example 1; SEQ ID NO 5978; 243pp; English.

XX The invention relates to an isolated nucleic acid derived from  
 CC Enterococcus faecium encoding an Enterococcus faecium polypeptide having  
 CC one of 10 fully defined sequences given in the (or comprising 40  
 CC sequential nucleotides chosen from any of the nucleic acids, its  
 CC complement or sequences hybridising to it). Also included are a  
 CC recombinant vector comprising the nucleic acid operably linked to  
 CC transcription regulatory element, a cell comprising the vector and a  
 CC single-stranded probe comprising the nucleic acid. The nucleic acids are  
 CC chosen from 3654 disclosed sequences encoding 3654 disclosed proteins.  
 CC The nucleic acids are useful for diagnosing pathological conditions  
 CC resulting from E. faecium bacterial infection (e.g. urinary tract  
 CC infection), bacteraemia, endocarditis, wounds and abdominal-pelvic  
 CC infection) and for screening drugs such as agonists and antagonists. The  
 CC nucleic acid is useful for recombinant production of Candida albicans -  
 CC derived peptides or antisense polypeptides. Pharmaceutical compositions  
 CC and vaccines containing the nucleic acid are useful for preventing or  
 CC treating Enterococcus faecium infections. The present sequence represents  
 CC one if the disclosed E. faecium proteins.

XX Sequence 204 AA;

Query Match 2.9%; Score 7; DB 7; Length 204;  
 Best Local Similarity 100.0%; Pred. No. 5.8e+02;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 33 SGLASAP 39  
 |||||  
 DB 98 SGLASAP 104

RESULT 59  
 AAY35431  
 ID AAY35431 standard; protein; 228 AA.

XX AC AAY35431;  
 XX  
 DT 17-OCT-2003 (revised)  
 DT 13-SEP-1999 (first entry)  
 XX  
 DE Chlamydia pneumoniae transmembrane protein sequence.

XX KW Respiratory disease; pneumonia; bronchitis; heart disease; sarcoidosis;  
 KW sinusitis; purulent otitis media; erythema nodosum; pharyngitis; vaccine;  
 KW neutralising epitope.

XX OS Chlamydomphila pneumoniae.  
 XX PN WO9927105-A2.  
 XX PD 03-JUN-1999.  
 XX PF 20-NOV-1998; 98WO-IB001890.  
 XX PR 21-NOV-1997; 97FR-00014673.  
 XX PR 04-NOV-1998; 98US-0107078P.

XX (GIST ) GENSET.  
 XX PI Griffais R;  
 XX DR WPI; 1999-357842/30.  
 XX  
 XX PT Genome sequence of Chlamydia pneumoniae.  
 XX PS Page 1212; Disclosure; 1912pp; English.

XX CC AAY34584-Y35879 represent the proteins encoded by all the open reading  
 XX CC frames in the complete genome (see AAX91990) of Chlamydia pneumoniae. C.  
 XX CC pneumoniae causes respiratory disease such as pneumonia and bronchitis  
 XX CC and is thought to be a contributing factor in heart disease, sarcoidosis,  
 XX CC sinusitis, purulent otitis media, erythema nodosum or pharyngitis. The  
 XX CC polypeptides encoded by the open reading frames of the C. pneumoniae



CC genome (see AAY34584-Y35879) can be used in immunogenic compositions as  
 CC vaccines. Vectors containing C. pneumoniae nucleotide sequences can also  
 CC be used as immunogenic compositions, especially where the vector directs  
 CC the expression of a neutralising epitope of C. pneumoniae. (Updated on 17  
 CC -OCT-2003 to standardise OS field)

XX SQ Sequence 228 AA;

Query Match 2.9%; Score 7; DB 2; Length 228;  
 Best Local Similarity 100.0%; Pred. No. 6.4e-02;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 143 KDLGTLG 149  
 |||||  
 Db 158 KDLGTLG 164

RESULT 60

AAU17162

ID AAU17162 standard; protein; 237 AA.

XX AC AAU17162;

XX DT 07-NOV-2001 (first entry)

XX DE Novel signal transduction pathway protein, Seq ID 727.

XX KW Neuroprotective; cytostatic; dermatological; immunosuppressive; tumour;  
 KW antiinflammatory; anti-HIV; antibacterial; antiinflammatory; cancer;  
 KW immune system disorder; rheumatoid arthritis; inflammatory condition;  
 KW organ transplant rejection; infection; hepatitis C; blood disorder;  
 KW sickle cell anaemia; hyperproliferative disorder; Gaucher's disease;  
 KW neurodegenerative disorder; Alzheimer's disease; Parkinson's disease;  
 KW chromosomal abnormality; Down syndrome; ischaemia; renal disorder;  
 KW cardiovascular; respiratory; wound healing; endocrine; Addison's disease;  
 KW reproductive system; gastrointestinal; liver disorder; AIDS;  
 KW acquired immune deficiency syndrome.

XX OS Homo sapiens.

XX PN WO200154733-A1.

XX PD 02-AUG-2001.

XX PF 17-JAN-2001; 2001WO-US001312.

XX PR 31-JAN-2000; 2000US-0179065P.

XX PR 04-FEB-2000; 2000US-0180628P.

XX PR 24-FEB-2000; 2000US-0184664P.

XX PR 02-MAR-2000; 2000US-0186350P.

XX PR 16-MAR-2000; 2000US-0189874P.

XX PR 17-MAR-2000; 2000US-0190076P.

XX PR 18-APR-2000; 2000US-0198123P.

XX PR 19-MAY-2000; 2000US-0205151P.

XX PR 07-JUN-2000; 2000US-0209467P.

XX PR 30-JUN-2000; 2000US-0215135P.

XX PR 07-JUL-2000; 2000US-0216647P.

XX PR 07-JUL-2000; 2000US-0216880P.

XX PR 11-JUL-2000; 2000US-0217487P.

XX PR 14-JUL-2000; 2000US-0217496P.

XX PR 26-JUL-2000; 2000US-0218290P.

XX PR 26-JUL-2000; 2000US-0220963P.

XX PR 14-AUG-2000; 2000US-0224518P.

XX PR 14-AUG-2000; 2000US-0224519P.

XX PR 14-AUG-2000; 2000US-0225213P.

PR 14-AUG-2000; 2000US-0225757P.  
 PR 14-AUG-2000; 2000US-0225758P.  
 PR 14-AUG-2000; 2000US-0225759P.  
 PR 18-AUG-2000; 2000US-0226279P.  
 PR 22-AUG-2000; 2000US-0226681P.  
 PR 22-AUG-2000; 2000US-0226688P.  
 PR 22-AUG-2000; 2000US-0227182P.  
 PR 23-AUG-2000; 2000US-0227009P.  
 PR 30-AUG-2000; 2000US-0228924P.  
 PR 01-SEP-2000; 2000US-0229287P.  
 PR 01-SEP-2000; 2000US-0229343P.  
 PR 01-SEP-2000; 2000US-0229344P.  
 PR 01-SEP-2000; 2000US-0229345P.  
 PR 05-SEP-2000; 2000US-0229509P.  
 PR 05-SEP-2000; 2000US-0229513P.  
 PR 06-SEP-2000; 2000US-0230437P.  
 PR 06-SEP-2000; 2000US-0230438P.  
 PR 08-SEP-2000; 2000US-0231242P.  
 PR 08-SEP-2000; 2000US-0231243P.  
 PR 08-SEP-2000; 2000US-0231244P.  
 PR 08-SEP-2000; 2000US-0231413P.  
 PR 08-SEP-2000; 2000US-0231414P.  
 PR 08-SEP-2000; 2000US-0232080P.  
 PR 08-SEP-2000; 2000US-0232081P.  
 PR 12-SEP-2000; 2000US-0231968P.  
 PR 14-SEP-2000; 2000US-0232397P.  
 PR 14-SEP-2000; 2000US-0232398P.  
 PR 14-SEP-2000; 2000US-0232399P.  
 PR 14-SEP-2000; 2000US-0232400P.  
 PR 14-SEP-2000; 2000US-0232401P.  
 PR 14-SEP-2000; 2000US-0233063P.  
 PR 14-SEP-2000; 2000US-0233065P.  
 PR 21-SEP-2000; 2000US-0234223P.  
 PR 21-SEP-2000; 2000US-0234274P.  
 PR 25-SEP-2000; 2000US-0234997P.  
 PR 25-SEP-2000; 2000US-0234998P.  
 PR 26-SEP-2000; 2000US-0235484P.  
 PR 27-SEP-2000; 2000US-0235834P.  
 PR 27-SEP-2000; 2000US-0235836P.  
 PR 29-SEP-2000; 2000US-0236327P.  
 PR 29-SEP-2000; 2000US-0236367P.  
 PR 29-SEP-2000; 2000US-0236368P.  
 PR 29-SEP-2000; 2000US-0236369P.  
 PR 29-SEP-2000; 2000US-0236370P.  
 PR 02-OCT-2000; 2000US-0236802P.  
 PR 02-OCT-2000; 2000US-0237037P.  
 PR 02-OCT-2000; 2000US-0237038P.  
 PR 02-OCT-2000; 2000US-0237039P.  
 PR 02-OCT-2000; 2000US-0237040P.  
 PR 13-OCT-2000; 2000US-0239935P.  
 PR 13-OCT-2000; 2000US-0239937P.  
 PR 20-OCT-2000; 2000US-0240960P.  
 PR 20-OCT-2000; 2000US-0241221P.  
 PR 20-OCT-2000; 2000US-0241785P.  
 PR 20-OCT-2000; 2000US-0241786P.  
 PR 20-OCT-2000; 2000US-0241787P.  
 PR 20-OCT-2000; 2000US-0241808P.  
 PR 20-OCT-2000; 2000US-0241809P.  
 PR 20-OCT-2000; 2000US-0241828P.  
 PR 01-NOV-2000; 2000US-0244617P.  
 PR 08-NOV-2000; 2000US-0246474P.  
 PR 08-NOV-2000; 2000US-0246475P.  
 PR 08-NOV-2000; 2000US-0246476P.  
 PR 08-NOV-2000; 2000US-0246477P.  
 PR 08-NOV-2000; 2000US-0246478P.  
 PR 08-NOV-2000; 2000US-0246523P.  
 PR 08-NOV-2000; 2000US-0246524P.  
 PR 08-NOV-2000; 2000US-0246525P.  
 PR 08-NOV-2000; 2000US-0246526P.  
 PR 08-NOV-2000; 2000US-0246527P.  
 PR 08-NOV-2000; 2000US-0246528P.  
 PR 08-NOV-2000; 2000US-0246532P.

PR 08-NOV-2000; 2000US-0246609P.  
 PR 08-NOV-2000; 2000US-0246610P.  
 PR 08-NOV-2000; 2000US-0246611P.  
 PR 08-NOV-2000; 2000US-0246613P.  
 PR 17-NOV-2000; 2000US-0249207P.  
 PR 17-NOV-2000; 2000US-0249208P.  
 PR 17-NOV-2000; 2000US-0249209P.  
 PR 17-NOV-2000; 2000US-0249210P.  
 PR 17-NOV-2000; 2000US-0249211P.  
 PR 17-NOV-2000; 2000US-0249212P.  
 PR 17-NOV-2000; 2000US-0249213P.  
 PR 17-NOV-2000; 2000US-0249214P.  
 PR 17-NOV-2000; 2000US-0249215P.  
 PR 17-NOV-2000; 2000US-0249216P.  
 PR 17-NOV-2000; 2000US-0249217P.  
 PR 17-NOV-2000; 2000US-0249218P.  
 PR 17-NOV-2000; 2000US-0249244P.  
 PR 17-NOV-2000; 2000US-0249245P.  
 PR 17-NOV-2000; 2000US-0249246P.  
 PR 17-NOV-2000; 2000US-0249265P.  
 PR 17-NOV-2000; 2000US-0249297P.  
 PR 17-NOV-2000; 2000US-0249299P.  
 PR 01-DEC-2000; 2000US-0249300P.  
 PR 01-DEC-2000; 2000US-0250160P.  
 PR 01-DEC-2000; 2000US-0250391P.  
 PR 05-DEC-2000; 2000US-0251030P.  
 PR 05-DEC-2000; 2000US-0251988P.  
 PR 05-DEC-2000; 2000US-0256719P.  
 PR 06-DEC-2000; 2000US-0251479P.  
 PR 08-DEC-2000; 2000US-0251856P.  
 PR 08-DEC-2000; 2000US-0251868P.  
 PR 08-DEC-2000; 2000US-0251869P.  
 PR 08-DEC-2000; 2000US-0251989P.  
 PR 08-DEC-2000; 2000US-0251990P.  
 PR 11-DEC-2000; 2000US-0254907P.  
 PR 05-JAN-2001; 2001US-0255967P.  
 (HUMA-) HUMAN GENOME SCI INC.

PI Rosen CA, Barash SC, Ruben SM;  
 DR WPI; 2001-465460/50.  
 DR N-PSDB; AAS27079.

PT Novel polypeptides useful for diagnosing, treating, preventing and/or  
 PT prognosing disorders related to the proteins, including cancers, immune  
 PT disorders and neuronal disorders.

PS Claim 1; SEQ ID NO 727; 880pp; English.

XX The invention relates to novel isolated polypeptides (I), and  
 CC polynucleotides (II). (I), (II) and the antibody to (I) are useful for  
 CC diagnosing, preventing and treating diseases including immune system  
 CC disorders (e.g. congenital and acquired immunodeficiencies, autoimmune  
 CC disorders (e.g. rheumatoid arthritis), inflammatory conditions, organ  
 CC transplant rejections and graft versus host disease, infectious diseases  
 CC (e.g. hepatitis C), bleeding disorders, haemoglobin abnormalities and  
 CC other blood-related disorders (sickle cell anaemia), myeloproliferative  
 CC disorders, primary haematopoietic disorders, hyperproliferative disorders  
 CC (e.g. Gaucher's disease and cancer), neurodegenerative disorders (e.g.  
 CC Alzheimer's disease, Parkinson's disease), chromosomal abnormalities  
 CC (Down syndrome), ischaemic injury (e.g. stroke), renal disorders (e.g.  
 CC glomerulonephritis), cardiovascular disorders (e.g. arrhythmia),  
 CC respiratory disorders, dermatological disorders, in wound healing,  
 CC epithelial cell proliferation, endocrine disorders (e.g. Addison's  
 CC disease), reproductive system disorders, gastrointestinal disorder  
 CC (inflammatory disorders), liver disorders (cirrhosis), as stimulators of  
 CC B-cell responsiveness to pathogens, activators of T-cells, to induce  
 CC higher affinity antibodies, and as a means to induce tumour proliferation  
 CC in pathologies e.g. acquired immune deficiency syndrome (AIDS). AAU17059-  
 CC AAU17683 represent novel signal transduction pathway protein, amino acid  
 CC sequences of the invention

Query Match 2.9%; Score 7; DB 4; Length 237;  
 Best Local Similarity 100.0%; Pred. No. 6.6e+02;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 128 VIGISQR 134  
 DB 215 VIG-SQR 221  
 RESULT 61  
 AAM84513  
 ID AAM84513 standard; protein; 237 AA.  
 AC AAM84513;  
 XX  
 XX 07-NOV-2001 (first entry)  
 XX Human immune/haematopoietic antigen SEQ ID NO:12106.  
 XX Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;  
 KW cytostatic; gene therapy; vaccine; metastasis.  
 XX Homo sapiens.  
 XX WO200157182-A2.  
 XX 09-AUG-2001.  
 XX 17-JAN-2001; 2001WO-US001354.  
 XX 31-JAN-2000; 2000US-0179065P.  
 PR 04-FEB-2000; 2000US-0180628P.  
 PR 24-FEB-2000; 2000US-0184664P.  
 PR 02-MAR-2000; 2000US-0186350P.  
 PR 16-MAR-2000; 2000US-0189874P.  
 PR 17-MAR-2000; 2000US-0190076P.  
 PR 18-APR-2000; 2000US-0198123P.  
 PR 19-MAY-2000; 2000US-0205515P.  
 PR 07-JUN-2000; 2000US-0209467P.  
 PR 28-JUN-2000; 2000US-0214886P.  
 PR 30-JUN-2000; 2000US-0215135P.  
 PR 07-JUL-2000; 2000US-0216647P.  
 PR 07-JUL-2000; 2000US-0216880P.  
 PR 11-JUL-2000; 2000US-0217487P.  
 PR 11-JUL-2000; 2000US-0217496P.  
 PR 14-JUL-2000; 2000US-0218290P.  
 PR 26-JUL-2000; 2000US-0220963P.  
 PR 26-JUL-2000; 2000US-0220964P.  
 PR 14-AUG-2000; 2000US-0224518P.  
 PR 14-AUG-2000; 2000US-0224519P.  
 PR 14-AUG-2000; 2000US-0225213P.  
 PR 14-AUG-2000; 2000US-0225214P.  
 PR 14-AUG-2000; 2000US-0225266P.  
 PR 14-AUG-2000; 2000US-0225267P.  
 PR 14-AUG-2000; 2000US-0225268P.  
 PR 14-AUG-2000; 2000US-0225270P.  
 PR 14-AUG-2000; 2000US-0225447P.  
 PR 14-AUG-2000; 2000US-0225757P.  
 PR 14-AUG-2000; 2000US-0225758P.  
 PR 14-AUG-2000; 2000US-0225759P.  
 PR 14-AUG-2000; 2000US-0226279P.  
 PR 14-AUG-2000; 2000US-0226681P.  
 PR 22-AUG-2000; 2000US-0226988P.  
 PR 22-AUG-2000; 2000US-0227182P.  
 PR 23-AUG-2000; 2000US-0227009P.  
 PR 30-AUG-2000; 2000US-0228924P.  
 PR 01-SEP-2000; 2000US-0229287P.  
 PR 01-SEP-2000; 2000US-0229343P.  
 PR 01-SEP-2000; 2000US-0229344P.  
 PR 01-SEP-2000; 2000US-0229345P.  
 PR 05-SEP-2000; 2000US-0229509P.  
 PR 05-SEP-2000; 2000US-0229513P.



KW Central nervous system; CNS; autoimmune disease; rheumatoid arthritis;  
 KW hyperproliferative disorder; neoplasm; cardiovascular disease;  
 KW cardiac arrest; cerebrovascular disorder; ischaemia; angiogenesis;  
 KW nervous system disorder; Alzheimer's disease; AIDS; ocular disorder;  
 KW acquired immunodeficiency virus; dysphagia; gastrointestinal disorder;  
 KW adenocarcinoma; reproductive system disorder; testicular feminisation;  
 KW endocrine disorder; diabetes; cancer; leukaemia; neovascularisation;  
 KW respiratory disorder; renal disorder; kidney failure; blood disorder;  
 KW myocardial infarction; wound healing; cell proliferation; skin aging;  
 KW food additive; food preservative; gene therapy.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200155318-A2.  
 XX  
 XX 02-AUG-2001.  
 XX  
 XX 17-JAN-2001; 2001WO-US001332.  
 XX  
 PR 31-JAN-2000; 2000US-0179065P.  
 PR 04-FEB-2000; 2000US-0180628P.  
 PR 24-FEB-2000; 2000US-0184664P.  
 PR 02-MAR-2000; 2000US-0186350P.  
 PR 16-MAR-2000; 2000US-0189874P.  
 PR 17-MAR-2000; 2000US-0190076P.  
 PR 18-APR-2000; 2000US-0198123P.  
 PR 19-MAY-2000; 2000US-0205515P.  
 PR 07-JUN-2000; 2000US-0209467P.  
 PR 28-JUN-2000; 2000US-0214886P.  
 PR 30-JUN-2000; 2000US-0215135P.  
 PR 07-JUL-2000; 2000US-0216647P.  
 PR 07-JUL-2000; 2000US-0216880P.  
 PR 11-JUL-2000; 2000US-0217487P.  
 PR 11-JUL-2000; 2000US-0217496P.  
 PR 14-JUL-2000; 2000US-0218290P.  
 PR 26-JUL-2000; 2000US-0220963P.  
 PR 26-JUL-2000; 2000US-0220964P.  
 PR 14-AUG-2000; 2000US-0224518P.  
 PR 14-AUG-2000; 2000US-0224519P.  
 PR 14-AUG-2000; 2000US-0225213P.  
 PR 14-AUG-2000; 2000US-0225214P.  
 PR 14-AUG-2000; 2000US-0225466P.  
 PR 14-AUG-2000; 2000US-0225467P.  
 PR 14-AUG-2000; 2000US-0225468P.  
 PR 14-AUG-2000; 2000US-0225270P.  
 PR 14-AUG-2000; 2000US-0225447P.  
 PR 14-AUG-2000; 2000US-0225757P.  
 PR 14-AUG-2000; 2000US-0225758P.  
 PR 14-AUG-2000; 2000US-0225759P.  
 PR 18-AUG-2000; 2000US-0226279P.  
 PR 22-AUG-2000; 2000US-0226681P.  
 PR 22-AUG-2000; 2000US-0226688P.  
 PR 23-AUG-2000; 2000US-0227182P.  
 PR 23-AUG-2000; 2000US-0227009P.  
 PR 30-AUG-2000; 2000US-0228924P.  
 PR 01-SEP-2000; 2000US-0229257P.  
 PR 01-SEP-2000; 2000US-0229343P.  
 PR 01-SEP-2000; 2000US-0229344P.  
 PR 01-SEP-2000; 2000US-0229345P.  
 PR 05-SEP-2000; 2000US-0229509P.  
 PR 05-SEP-2000; 2000US-0229513P.  
 PR 06-SEP-2000; 2000US-0230437P.  
 PR 06-SEP-2000; 2000US-0230438P.  
 PR 08-SEP-2000; 2000US-0231242P.  
 PR 08-SEP-2000; 2000US-0231243P.  
 PR 08-SEP-2000; 2000US-0231244P.  
 PR 08-SEP-2000; 2000US-0231413P.  
 PR 08-SEP-2000; 2000US-0231414P.  
 PR 08-SEP-2000; 2000US-0232080P.  
 PR 08-SEP-2000; 2000US-0232081P.  
 PR 12-SEP-2000; 2000US-0231968P.  
 PR 14-SEP-2000; 2000US-0232397P.  
 PR 14-SEP-2000; 2000US-0232398P.  
 PR 14-SEP-2000; 2000US-0232399P.  
 PR 14-SEP-2000; 2000US-0232400P.  
 PR 14-SEP-2000; 2000US-0232401P.  
 PR 14-SEP-2000; 2000US-0233063P.  
 PR 14-SEP-2000; 2000US-0233064P.  
 PR 14-SEP-2000; 2000US-0233065P.  
 PR 21-SEP-2000; 2000US-0233223P.  
 PR 21-SEP-2000; 2000US-0234274P.  
 PR 25-SEP-2000; 2000US-0234997P.  
 PR 25-SEP-2000; 2000US-0234998P.  
 PR 26-SEP-2000; 2000US-0235484P.  
 PR 27-SEP-2000; 2000US-0235834P.  
 PR 27-SEP-2000; 2000US-0235836P.  
 PR 29-SEP-2000; 2000US-0236327P.  
 PR 29-SEP-2000; 2000US-0236367P.  
 PR 29-SEP-2000; 2000US-0236368P.  
 PR 29-SEP-2000; 2000US-0236369P.  
 PR 29-SEP-2000; 2000US-0236370P.  
 PR 02-OCT-2000; 2000US-0236802P.  
 PR 02-OCT-2000; 2000US-0237037P.  
 PR 02-OCT-2000; 2000US-0237038P.  
 PR 02-OCT-2000; 2000US-0237039P.  
 PR 12-OCT-2000; 2000US-0237040P.  
 PR 13-OCT-2000; 2000US-0239935P.  
 PR 13-OCT-2000; 2000US-0239937P.  
 PR 20-OCT-2000; 2000US-0240960P.  
 PR 20-OCT-2000; 2000US-0241221P.  
 PR 20-OCT-2000; 2000US-0241785P.  
 PR 20-OCT-2000; 2000US-0241786P.  
 PR 20-OCT-2000; 2000US-0241787P.  
 PR 20-OCT-2000; 2000US-0241808P.  
 PR 20-OCT-2000; 2000US-0241809P.  
 PR 20-OCT-2000; 2000US-0241826P.  
 PR 01-NOV-2000; 2000US-0244617P.  
 PR 08-NOV-2000; 2000US-0246474P.  
 PR 08-NOV-2000; 2000US-0246475P.  
 PR 08-NOV-2000; 2000US-0246476P.  
 PR 08-NOV-2000; 2000US-0246477P.  
 PR 08-NOV-2000; 2000US-0246478P.  
 PR 08-NOV-2000; 2000US-0246523P.  
 PR 08-NOV-2000; 2000US-0246524P.  
 PR 08-NOV-2000; 2000US-0246525P.  
 PR 08-NOV-2000; 2000US-0246526P.  
 PR 08-NOV-2000; 2000US-0246527P.  
 PR 08-NOV-2000; 2000US-0246528P.  
 PR 08-NOV-2000; 2000US-0246532P.  
 PR 08-NOV-2000; 2000US-0246609P.  
 PR 08-NOV-2000; 2000US-0246610P.  
 PR 08-NOV-2000; 2000US-0246611P.  
 PR 08-NOV-2000; 2000US-0246613P.  
 PR 17-NOV-2000; 2000US-0249207P.  
 PR 17-NOV-2000; 2000US-0249208P.  
 PR 17-NOV-2000; 2000US-0249209P.  
 PR 17-NOV-2000; 2000US-0249210P.  
 PR 17-NOV-2000; 2000US-0249211P.  
 PR 17-NOV-2000; 2000US-0249212P.  
 PR 17-NOV-2000; 2000US-0249213P.  
 PR 17-NOV-2000; 2000US-0249214P.  
 PR 17-NOV-2000; 2000US-0249215P.  
 PR 17-NOV-2000; 2000US-0249216P.  
 PR 17-NOV-2000; 2000US-0249217P.  
 PR 17-NOV-2000; 2000US-0249218P.  
 PR 17-NOV-2000; 2000US-0249244P.  
 PR 17-NOV-2000; 2000US-0249245P.  
 PR 17-NOV-2000; 2000US-0249246P.  
 PR 17-NOV-2000; 2000US-0249265P.  
 PR 17-NOV-2000; 2000US-0249297P.  
 PR 17-NOV-2000; 2000US-0249299P.  
 PR 17-NOV-2000; 2000US-0249300P.  
 PR 01-DEC-2000; 2000US-0250160P.  
 PR 01-DEC-2000; 2000US-0250391P.  
 PR 05-DEC-2000; 2000US-0251030P.  
 PR 05-DEC-2000; 2000US-0251988P.

Central nervous system, CNS; autoimmune disease, rheumatoid arthritis, hyperproliferative disorder, neoplasia, cardiovascular disorder  
cardiac arrest, cerebrovascular disorder, schizophrenia  
nervous system disorder, autoimmune disease, AIDS, drug overdose  
acquired immunodeficiency virus, dysphasia, gastrointestinal disorder, adenocarcinoma, reproductive system disorder, cervical ectometastasis, endocrine disorder, diabetes; cancer, leukaemia; neovascularisation;

[illegible]

PR 21-SEP-2000; 2000US-0234274P.  
PR 25-SEP-2000; 2000US-0234997P.  
PR 25-SEP-2000; 2000US-0234998P.  
PR 26-SEP-2000; 2000US-0235484P.  
PR 27-SEP-2000; 2000US-0235834P.  
PR 27-SEP-2000; 2000US-0235835P.  
PR 29-SEP-2000; 2000US-0236327P.  
PR 29-SEP-2000; 2000US-0236367P.  
PR 29-SEP-2000; 2000US-0236369P.  
PR 29-SEP-2000; 2000US-0236369P.  
PR 29-SEP-2000; 2000US-0236370P.  
PR 02-OCT-2000; 2000US-0236802P.  
PR 02-OCT-2000; 2000US-0237037P.  
PR 02-OCT-2000; 2000US-0237038P.  
PR 02-OCT-2000; 2000US-0237039P.  
PR 13-OCT-2000; 2000US-0237040P.  
PR 13-OCT-2000; 2000US-0239935P.  
PR 20-OCT-2000; 2000US-0241785P.  
PR 20-OCT-2000; 2000US-0241786P.  
PR 20-OCT-2000; 2000US-0241787P.  
PR 20-OCT-2000; 2000US-0241808P.  
PR 20-OCT-2000; 2000US-0241809P.  
PR 20-OCT-2000; 2000US-0241828P.  
PR 01-NOV-2000; 2000US-0244611P.  
PR 08-NOV-2000; 2000US-0246474P.  
PR 08-NOV-2000; 2000US-0246475P.  
PR 08-NOV-2000; 2000US-0246476P.  
PR 08-NOV-2000; 2000US-0246477P.  
PR 08-NOV-2000; 2000US-0246478P.  
PR 08-NOV-2000; 2000US-0246523P.  
PR 08-NOV-2000; 2000US-0246524P.  
PR 08-NOV-2000; 2000US-0246525P.  
PR 08-NOV-2000; 2000US-0246526P.  
PR 08-NOV-2000; 2000US-0246527P.  
PR 08-NOV-2000; 2000US-0246528P.  
PR 08-NOV-2000; 2000US-0246532P.  
PR 08-NOV-2000; 2000US-0246532P.  
PR 08-NOV-2000; 2000US-0246532P.  
PR 08-NOV-2000; 2000US-0246610P.  
PR 08-NOV-2000; 2000US-0246611P.  
PR 08-NOV-2000; 2000US-0246613P.  
PR 17-NOV-2000; 2000US-0249207P.  
PR 17-NOV-2000; 2000US-0249208P.  
PR 17-NOV-2000; 2000US-0249209P.  
PR 17-NOV-2000; 2000US-0249210P.  
PR 17-NOV-2000; 2000US-0249211P.  
PR 17-NOV-2000; 2000US-0249212P.  
PR 17-NOV-2000; 2000US-0249213P.  
PR 17-NOV-2000; 2000US-0249214P.  
PR 17-NOV-2000; 2000US-0249215P.  
PR 17-NOV-2000; 2000US-0249216P.  
PR 17-NOV-2000; 2000US-0249217P.  
PR 17-NOV-2000; 2000US-0249218P.  
PR 17-NOV-2000; 2000US-0249244P.  
PR 17-NOV-2000; 2000US-0249245P.  
PR 17-NOV-2000; 2000US-0249245P.  
PR 17-NOV-2000; 2000US-0249264P.  
PR 17-NOV-2000; 2000US-0249265P.  
PR 17-NOV-2000; 2000US-0249297P.  
PR 17-NOV-2000; 2000US-0249299P.  
PR 17-NOV-2000; 2000US-0249299P.  
PR 01-DEC-2000; 2000US-0250160P.  
PR 01-DEC-2000; 2000US-0250319P.  
PR 05-DEC-2000; 2000US-0251030P.  
PR 05-DEC-2000; 2000US-0251988P.  
PR 06-DEC-2000; 2000US-0256719P.  
PR 06-DEC-2000; 2000US-0251479P.  
PR 08-DEC-2000; 2000US-0251856P.  
PR 08-DEC-2000; 2000US-0251868P.  
PR 08-DEC-2000; 2000US-0251869P.  
PR 08-DEC-2000; 2000US-0251989P.  
PR 08-DEC-2000; 2000US-0251990P.

PR 11-DEC-2000; 2000US-0254097P.  
PR 05-JAN-2001; 2001US-0259678P.  
PR (HUMA-) HUMAN GENOME SCI INC.  
XX Rosen CA, Barash SC, Ruben SM;  
XX WPI; 2001-581633/65.  
XX N-PSDB; ABK43703.  
PT New isolated nucleic acid encoding a protein for diagnosing, preventing,  
PT treating or ameliorating medical conditions and used as food additives or  
PT preservatives.  
XX Claim 9; SEQ ID NO 891; 837pp; English.  
XX The invention describes an isolated nucleic acid molecule (I) encoding a  
CC novel central nervous system protein. (I) and polypeptides (III) encoded  
CC by (I), are used to treat a medical conditions and in diagnosis of a  
CC pathological condition. Disorders which are diagnosed or treated include  
CC autoimmune diseases e.g. rheumatoid arthritis, hyperproliferative  
CC disorders e.g. neoplasms of the breast or liver, cardiovascular disorders  
CC e.g. cardiac arrest, cerebrovascular disorders e.g. cerebral ischaemia,  
CC angiogenesis, nervous system disorders e.g. Alzheimer's disease and  
CC amyotrophic lateral sclerosis, infections caused by bacteria, viruses  
CC e.g. Acquired immunodeficiency virus (AIDS) and fungi, ocular disorders  
CC e.g. corneal infection, gastrointestinal disorders e.g. dysphagia,  
CC adenocarcinomas and irritable bowel syndrome, reproductive system  
CC disorders e.g. testicular feminisation, endocrine disorders e.g. diabetes  
CC and pituitary dwarfism, cancers and disorders at the cellular level e.g.  
CC leukaemia, disorders involving neovascularisation e.g. malignancies,  
CC respiratory disorders e.g. nonallergic rhinitis, renal disorders e.g.  
CC acute kidney failure and blood related disorders e.g. myocardial  
CC infarction. The polypeptides can also be used to aid wound healing and  
CC epithelial cell proliferation, to prevent skin aging due to sunburn, to  
CC maintain organs before transplantation, for supporting cell culture of  
CC primary tissues, to regenerate tissues and in chemotaxis. The  
CC polypeptides can also be used as a food additive or preservative to  
CC increase or decrease storage capabilities, fat content, lipid, protein,

Query Match 2.9%; Score 7; DB 4; Length 237;

Best Local Similarity 100.0%; Pred. No. 6.6e+02; Mismatches 0; Indels 0; Gaps 0;  
Matches 7; Conservative 0;

Qy 128 VIGISQR 134  
| | | | |  
Db 215 VIGISQR 221

RESULT 64

ADB93870

ID ADB93870 standard; protein; 237 AA.

XX ADB93870;

XX DT 04-DEC-2003 (first entry)

XX DE Human novel protein #104.

XX human; autoimmune disease; Parkinson's disease; silicosis;  
KW gastrointestinal disease; atherosclerosis; haemophilia; thrombocytopenia;  
KW immunosuppressive agent; adjuvant; enhance immune response;  
KW higher affinity antibody induction;  
KW increased serum immunoglobulin concentration.

XX OS Homo sapiens.

XX FN US2002168711-A1.

XX PD 14-NOV-2002.

XX PP 17-JAN-2001; 2001US-00764868.

XX

PR 31-JAN-2000; 2000US-0179065P.  
PR 04-FEB-2000; 2000US-0180628P.  
PR 28-JUN-2000; 2000US-0214886P.  
PR 07-JUL-2000; 2000US-0216647P.  
PR 07-JUL-2000; 2000US-0216880P.  
PR 11-JUL-2000; 2000US-0217480P.  
PR 11-JUL-2000; 2000US-0217496P.  
PR 14-JUL-2000; 2000US-0218290P.  
PR 26-JUL-2000; 2000US-0220963P.  
PR 26-JUL-2000; 2000US-0220964P.  
PR 14-AUG-2000; 2000US-0224518P.  
PR 14-AUG-2000; 2000US-0224519P.  
PR 14-AUG-2000; 2000US-0225267P.  
PR 14-AUG-2000; 2000US-0225268P.  
PR 14-AUG-2000; 2000US-0225270P.  
PR 14-AUG-2000; 2000US-0225447P.  
PR 14-AUG-2000; 2000US-0225757P.  
PR 14-AUG-2000; 2000US-0225758P.  
PR 22-AUG-2000; 2000US-0226868P.  
PR 30-AUG-2000; 2000US-0228924P.  
PR 01-SEP-2000; 2000US-0229287P.  
PR 01-SEP-2000; 2000US-0229343P.  
PR 01-SEP-2000; 2000US-0229344P.  
PR 01-SEP-2000; 2000US-0229345P.  
PR 05-SEP-2000; 2000US-0229509P.  
PR 05-SEP-2000; 2000US-0229513P.  
PR 08-SEP-2000; 2000US-0231413P.  
PR 21-SEP-2000; 2000US-0234272P.  
PR 21-SEP-2000; 2000US-0234273P.  
PR 25-SEP-2000; 2000US-0234977P.  
PR 27-SEP-2000; 2000US-0235834P.  
PR 29-SEP-2000; 2000US-0236327P.  
PR 29-SEP-2000; 2000US-0236367P.  
PR 29-SEP-2000; 2000US-0236368P.  
PR 29-SEP-2000; 2000US-0236369P.  
PR 29-SEP-2000; 2000US-0236370P.  
PR 02-OCT-2000; 2000US-0236802P.  
PR 02-OCT-2000; 2000US-0237037P.  
PR 02-OCT-2000; 2000US-0237038P.  
PR 02-OCT-2000; 2000US-0237039P.  
PR 13-OCT-2000; 2000US-0237040P.  
PR 20-OCT-2000; 2000US-0239935P.  
PR 20-OCT-2000; 2000US-0240960P.  
PR 20-OCT-2000; 2000US-0241785P.  
PR 20-OCT-2000; 2000US-0241809P.  
PR 01-NOV-2000; 2000US-0244617P.  
PR 17-NOV-2000; 2000US-0249299P.  
PR 08-DEC-2000; 2000US-0251856P.  
PR 08-DEC-2000; 2000US-0251868P.  
PR 08-DEC-2000; 2000US-0251869P.  
XX (ROSE/) ROSEN C A.  
PA (RUBE/) RUBEN S M.  
PA (BARA/) BARASH S C.  
XX  
PI Rosen CA, Ruben SM, Barash SC;  
XX  
XX WPI; 2003-719985/68.  
DR N-PSDB; ADB93247.  
XX  
XX New isolated polypeptide useful for diagnosing and treating  
PT immunosuppressive conditions such as autoimmune disease and Parkinson's  
PT disease.  
XX  
XX Claim 11; SEQ ID NO 727; 345pp; English.  
XX  
XX The invention relates to an isolated polypeptide. The polypeptide is  
CC useful for diagnosing a pathological condition or a susceptibility to a  
CC pathological condition in a subject, by determining the presence or  
CC amount of expression of the polypeptide in a biological sample and  
CC diagnosing a pathological condition or a susceptibility to a pathological  
CC condition based on the presence or amount of expression of the  
CC polypeptide. The polypeptide is also useful for identifying a binding

CC partner to the polypeptide, which involves contacting the polypeptide  
CC with a binding partner and determining whether the binding partner  
CC effects an activity of the polypeptide. The polypeptide or the nucleic  
CC acid encoding the polypeptide is useful for preventing, treating, or  
CC ameliorating a medical condition, which involves administering the  
CC polypeptide or the nucleic acid to a mammalian subject. The nucleic acid  
CC is useful for diagnosing a pathological condition or a susceptibility to  
CC a pathological condition in a subject, which involves determining the  
CC presence or absence of a mutation in the nucleic acid, and diagnosing a  
CC pathological condition or susceptibility to a pathological condition  
CC based on the presence or absence of the mutation. The polypeptide, the  
CC nucleic acid and an antibody to the polypeptide are useful for treating  
CC autoimmune disease, Parkinson's disease, silicosis, gastrointestinal  
CC disease, atherosclerosis, haemophilia, thrombocytopenia. The polypeptide,  
CC the nucleic acid and the antibody are useful as immunosuppressive agents,  
CC as adjuvants to enhance immune responses, and as agents to induce higher  
CC affinity antibodies and increase serum immunoglobulin concentrations. The  
CC present sequence represents the amino acid sequence of a novel human  
CC protein. Note: The sequence data for this patent did not form part of the  
CC printed specification but was obtained in electronic format direct from  
CC USPTO at seqdata.uspto.gov/sequence.html?DocID=20020168711.  
XX  
SQ Sequence 237 AA;  
  
Query Match 2.9%; Score 7; DB 7; Length 237;  
Best Local Similarity 100.0%; Pred. No. 6.6e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
Qy 128 VIGISQR 134  
Db 215 VIGISQR 221  
  
RESULT 65  
ABM70309  
ID ABM70309 standard; protein; 248 AA.  
AC ABM70309;  
XX  
XX 20-NOV-2003 (first entry)  
DE Photorhabdus luminescens protein sequence #3406.  
XX  
XX Antibacterial; fungicide; insecticide; polymorphism; genetic analysis;  
XX detection; food; gene expression; plant; animal; microorganism; toxin;  
XX antibiotic; biopesticide; virulence factor; disease model; plague;  
XX whooping cough.  
XX  
XX Photorhabdus luminescens.  
XX  
XX WO200294867-A2.  
XX  
XX 28-NOV-2002.  
XX  
XX 07-FEB-2002; 2002WO-IB003040.  
XX  
XX 07-FEB-2001; 2001FR-00001659.  
XX  
XX (INSP ) INST PASTEUR.  
XX (CNRS ) CNRS CENT NAT RECH SCI.  
XX  
XX Duchaud E, Taourit S, Glaser P, Frangeul L, Kunst F, Danchin A;  
PI Suchrieser C;  
XX  
XX WPI; 2003-148459/14.  
XX  
XX Genomic sequence of Photorhabdus luminescens and encoded polypeptides,  
XX useful e.g. as therapeutic antimicrobials and agricultural pesticides.  
XX  
XX Claim 2; SEQ ID NO 3406; 1205pp; French.  
XX  
XX The invention relates to the isolation of genes and their encoded  
CC proteins from Photorhabdus luminescens. The isolated sequences are

CC sources of probes and primers for detecting the genome of *P. luminescens*  
 CC and related species; to study polymorphisms; for gene analysis and for  
 CC detection/amplification of the genes. Antibodies (Ab) raised against the  
 CC polypeptides encoded by the genes are used for detection/identification  
 CC of *P. luminescens*, e.g. in foods. The genes, proteins, Ab and cells that  
 CC carry a gene-containing vector are used to select compounds that  
 CC modulate, regulate, induce or inhibit expression of the genes in plants,  
 CC animals or microorganisms other than *P. luminescens* and are able to alter  
 CC response or sensitivity to toxins and antibiotics produced by *P.*  
 CC *luminescens*. Cells transformed to express the genes are useful for  
 CC recombinant production of the proteins, particularly toxins and  
 CC antibacterials useful as insecticides, bactericides and fungicides. The  
 CC genes, proteins, vectors containing the genes and Ab are also useful  
 CC therapeutically (to treat microbial infection by bacteria or fungi that  
 CC are sensitive to *P. luminescens*-encoded toxins or antibiotics) and as  
 CC biopesticides. Other uses of the genes and the proteins are as virulence  
 CC factors and for identifying targets of human diseases for which *P.*  
 CC *luminescens* is a model (particularly plague and whooping cough). This  
 CC sequence represents one of the isolated *P. luminescens* proteins  
 XX  
 XX SQ Sequence 248 AA;

Query Match 2.9%; Score 7; DB 6; Length 248;  
 Best Local Similarity 100.0%; Pred. No. 6.9e+02; Indels 0; Gaps 0;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 160 IIAIGAG 166  
 |||||  
 Db 119 IIAIGAG 125

RESULT 66  
 AAU34184  
 ID AAU34184 standard; protein; 258 AA.

AC AAU34184;  
 DT 14-FEB-2002 (first entry)

DE Staphylococcus aureus cellular proliferation protein #450.

XX Antisense; prokaryotic cellular proliferation protein; antibiotic;  
 KW antibacterial; drug design.

XX Staphylococcus aureus.

XX WO200170955-A2.

XX 27-SEP-2001.

XX 21-MAR-2001; 2001WO-US009180.

XX 21-MAR-2000; 2000US-0191078P.

XX 23-MAY-2000; 2000US-0206848P.

XX 26-MAY-2000; 2000US-0207727P.

XX 23-OCT-2000; 2000US-0242578P.

XX 27-NOV-2000; 2000US-0253625P.

XX 22-DEC-2000; 2000US-0257931P.

XX 16-FEB-2001; 2001US-0269308P.

XX (ELIT-) ELITRA PHARM INC.

XX Haselbeck R, Ohlsen KL, Zyskind JW, Wall D, Trawick JD, Carr GJ;  
 PI Yamamoto RT, Xu HH;

XX WPI; 2001-611495/70.

XX N-PSDB; AAS52043.

XX New polynucleotides for the identification and development of  
 PT antibiotics, comprise sequences of antisense nucleic acids.

XX Example 3; SEQ ID NO 5680; 511bp; English.

CC The invention relates to antisense inhibitors of genes essential to  
 CC prokaryotic cellular proliferation, their use in identifying the genes,  
 CC their use in the discovery of novel antibiotics, the essential genes  
 CC themselves and the encoded proteins. The prokaryotes used are *Escherichia*  
 CC *coli*, *Staphylococcus aureus*, *Salmonella typhi*, *Klebsiella pneumoniae*,  
 CC *Pseudomonas aeruginosa* and *Enterococcus faecalis*. The invention is also  
 CC useful for the identification of potential new targets for antibiotic  
 CC development. The antisense nucleic acids can also be used to identify  
 CC proteins used in proliferation, to express these proteins, and to obtain  
 CC antibodies capable of binding to the expressed proteins. The proteins can  
 CC be used to screen compounds in rational drug discovery programmes. The  
 CC antisense nucleic acid sequence is also useful to screen for homologous  
 CC nucleic acids which are required for cell proliferation in a wide variety  
 CC of organisms. The present sequence represents an essential prokaryotic  
 CC cellular proliferation protein. Note: the sequence data for this patent  
 CC did not form part of the printed specification, but was obtained in  
 CC electronic format directly from WIPO at  
 CC ftp.wipo.int/pub/published\_pct\_sequences  
 XX  
 XX SQ Sequence 258 AA;

Query Match 2.9%; Score 7; DB 4; Length 258;  
 Best Local Similarity 100.0%; Pred. No. 7.1e+02;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 126 QPVIGIS 132  
 |||||  
 Db 56 QPVIGIS 62

RESULT 67  
 AAM41811  
 ID AAM41811 standard; protein; 272 AA.

XX AAM41811;

XX 22-OCT-2001 (first entry)

XX Human polypeptide SEQ ID NO 6742.

XX Human; neurotropic; immunosuppressant; cytostatic; gene therapy; cancer;  
 KW peripheral nervous system; neuropathy; central nervous system; CNS;  
 KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;  
 KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;  
 KW chemokinetic; thrombolytic; drug screening; arthritis; inflammation;  
 KW leukaemia.

XX Homo sapiens.

XX WO200153312-A1.

XX 26-JUL-2001.

XX 26-DEC-2000; 2000WO-US034263.

XX 23-DEC-1999; 98US-00471275.

XX 21-JAN-2000; 2000US-0048725.

XX 25-APR-2000; 2000US-00552317.

XX 20-JUN-2000; 2000US-00598042.

XX 19-JUL-2000; 2000US-00620312.

XX 03-AUG-2000; 2000US-00653450.

XX 14-SEP-2000; 2000US-00662191.

XX 19-OCT-2000; 2000US-00693036.

XX 29-NOV-2000; 2000US-00727344.

XX (HYSE-) HYSEQ INC.

XX Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;  
 PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J, Zhao QA;  
 PI Zhou P, Goodrich R, Drmanac RT;

XX WPI; 2001-442253/47.

XX N-PSDB; AAI60967.



XX Novel nucleic acids and polypeptides, useful for treating disorders such  
PT as central nervous system injuries.  
XX  
XX  
XX Example 2; SEQ ID NO 6742; 10078pp; English.  
XX  
XX The invention relates to human nucleic acids (AAI57798-AAI61369) and the  
CC encoded polypeptides (AAI38642-AAI42213) with neurotropic,  
CC immunosuppressant and cytoskeletal activity. The polynucleotides are useful  
CC in gene therapy. A composition containing a polypeptide or polynucleotide  
CC of the invention may be used to treat diseases of the peripheral nervous  
CC system, such as peripheral nervous injuries, peripheral neuropathy and  
CC localised neuropathies and central nervous system diseases, such as  
CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic  
CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the  
CC utilisation of the activities such as: Immune system suppression,  
CC Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic  
CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,  
CC assays for receptor activity, arthritis and inflammation, leukaemias and  
CC C.N.S disorders. Note: The sequence data for this patent did not form  
CC part of the printed specification  
XX  
XX Sequence 272 AA;

Query Match 2.9%; Score 7; DB 4; Length 272;  
Best Local Similarity 100.0%; Pred. No. 7.5e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 122 AAAVQPV 128  
Db 73 AAAVQPV 79

RESULT 68  
AAI41810  
ID AAI41810 standard; protein; 272 AA.

AC AAI41810;

XX 22-OCT-2001 (first entry)

XX Human polypeptide SEQ ID NO 6741.

XX Human; neurotropic; immunosuppressant; cytoskeletal; gene therapy; cancer;  
KW peripheral nervous system; neuropathy; central nervous system; CNS;  
KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;  
KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;  
KW chemokinetic; thrombolytic; drug screening; arthritis; inflammation;  
KW leukaemia.

OS Homo sapiens.

XX WO200153312-A1.

XX 26-JUL-2001.

XX 26-DEC-2000; 2000WO-US034263.

XX 23-DEC-1999; 99US-00471275.

XX 21-JAN-2000; 2000US-00488725.

XX 25-APR-2000; 2000US-0052317.

XX 20-JUN-2000; 2000US-00598042.

XX 19-JUL-2000; 2000US-00620312.

XX 03-AUG-2000; 2000US-00653450.

XX 14-SEP-2000; 2000US-00662191.

XX 19-OCT-2000; 2000US-00693036.

XX 29-NOV-2000; 2000US-00727344.

XX (HYSE-) HYSEQ INC.

XX Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;  
PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J, Zhao QA;  
PI Zhou P, Goodrich R, Drmanac RT;

XX WPI; 2001-442253/47.  
DR N-PSDB; AAI60966.  
XX Novel nucleic acids and polypeptides, useful for treating disorders such  
PT as central nervous system injuries.  
XX  
XX Example 2; SEQ ID NO 6741; 10078pp; English.  
XX  
XX The invention relates to human nucleic acids (AAI57798-AAI61369) and the  
CC encoded polypeptides (AAI38642-AAI42213) with neurotropic,  
CC immunosuppressant and cytoskeletal activity. The polynucleotides are useful  
CC in gene therapy. A composition containing a polypeptide or polynucleotide  
CC of the invention may be used to treat diseases of the peripheral nervous  
CC system, such as peripheral nervous injuries, peripheral neuropathy and  
CC localised neuropathies and central nervous system diseases, such as  
CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic  
CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the  
CC utilisation of the activities such as: Immune system suppression,  
CC Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic  
CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,  
CC assays for receptor activity, arthritis and inflammation, leukaemias and  
CC C.N.S disorders. Note: The sequence data for this patent did not form  
CC part of the printed specification  
XX  
XX Sequence 272 AA;

Query Match 2.9%; Score 7; DB 4; Length 272;  
Best Local Similarity 100.0%; Pred. No. 7.5e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 122 AAAVQPV 128  
Db 73 AAAVQPV 79

RESULT 69

ABBS9873

ID ABBS9873 standard; protein; 282 AA.

XX AC ABBS9873;

XX 26-MAR-2002 (first entry)

XX Drosophila melanogaster polypeptide SEQ ID NO 6411.

XX Drosophila; developmental biology; cell signalling; insecticide;  
KW Pharmaceutical.

XX Drosophila melanogaster.

XX WO200171042-A2.

XX 27-SEP-2001.

XX 23-MAR-2001; 2001WO-US009231.

XX 23-MAR-2000; 2000US-0191637P.

XX 11-JUL-2000; 2000US-00614150.

XX (PEXE) PE CORP NY.

XX Venter JC, Adams M, Li PWD, Myers EW;

XX WPI; 2001-656860/75.

XX N-PSDB; ABL03976.

XX New isolated nucleic acid detection reagent for detecting 1000 or more  
PT genes from Drosophila and for elucidating cell signalling and cell-cell  
PT interactions.

XX Disclosure; SEQ ID NO 6411; 21pp + Sequence Listing; English.

XX

CC The invention relates to an isolated nucleic acid detection reagent  
 CC capable of detecting 1000 or more genes from *Drosophila*. The invention is  
 CC useful in developmental biology and in elucidating cell signalling and  
 CC cell-cell interactions in higher eukaryotes for the development of  
 CC insecticides, therapeutics and pharmaceutical drugs. The invention  
 CC discloses genomic DNA sequences (ABU16176-ABU16051), expressed DNA  
 CC sequences (ABU01840-ABU16175) and the encoded proteins (ABU57737-  
 CC ABU72072). The sequence data for this patent did not form part of the  
 CC printed specification, but was obtained in electronic format directly  
 CC from WIPO at ftp.wipo.int/pub/published\_pct\_sequences  
 XX  
 XX Sequence 282 AA;

Query Match 2.9%; Score 7; DB 4; Length 282;  
 Best Local Similarity 100.0%; Pred. No. 7.7e+02;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 26 LNWLDAQ 32  
 |||||  
 DB 178 LNWLDAQ 184

RESULT 70  
 ABU16506  
 ID ABU16506 standard; protein; 286 AA.  
 XX  
 AC ABU16506;  
 XX  
 DT 19-JUN-2003 (first entry)  
 XX  
 DE Protein encoded by Prokaryotic essential gene #2033.  
 XX  
 KW Antisense; prokaryotic essential gene; cell proliferation; drug design.  
 XX  
 OS *Staphylococcus aureus*.  
 XX  
 PN WO200277183-A2.  
 XX  
 PD 03-OCT-2002.  
 XX  
 PF 21-MAR-2002; 2002WO-US009107.  
 XX  
 PR 21-MAR-2001; 2001US-00815242.  
 PR 06-SEP-2001; 2001US-00948993.  
 PR 25-OCT-2001; 2001US-0342923P.  
 PR 08-FEB-2002; 2002US-00072851.  
 PR 06-MAR-2002; 2002US-0362699P.  
 XX  
 PA (ELIT-) ELITRA PHARM INC.  
 XX  
 PI Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;  
 PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;  
 XX  
 DR WPI; 2003-029926/02.  
 DR N-PSDB; ACA20376.  
 XX  
 XX New antisense nucleic acids, useful for identifying proteins or screening  
 PT for homologous nucleic acids required for cellular proliferation to  
 PT isolate candidate molecules for rational drug discovery programs.  
 XX  
 PS Claim 25; SEQ ID NO 44430; 1766pp; English.

CC The invention relates to an isolated nucleic acid comprising any one of  
 CC the 6213 antisense sequences given in the specification where expression  
 CC of the nucleic acid inhibits proliferation of a cell. Also included are:  
 CC (1) a vector comprising a promoter operably linked to the nucleic acid  
 CC encoding a polypeptide whose expression is inhibited by the antisense  
 CC nucleic acid; (2) a host cell containing the vector; (3) an isolated  
 CC polypeptide or its fragment whose expression is inhibited by the  
 CC antisense nucleic acid; (4) an antibody capable of specifically binding  
 CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular  
 CC proliferation or the activity of a gene in an operon required for  
 CC proliferation; (7) identifying a compound that influences the activity of

CC the gene product or that has an activity against a biological pathway  
 CC required for proliferation, or that inhibits cellular proliferation; (8)  
 CC identifying a gene required for cellular proliferation or the biological  
 CC pathway in which a proliferation-required gene or its gene product lies  
 CC or a gene on which the test compound that inhibits proliferation of an  
 CC organism acts; (9) manufacturing an antibiotic; (10) profiling a  
 CC compound's activity; (11) a culture comprising strains in which the gene  
 CC product is overexpressed or underexpressed; (12) determining the extent  
 CC to which each of the strains is present in a culture or collection of  
 CC strains; or (13) identifying the target of a compound that inhibits the  
 CC proliferation of an organism. The antisense nucleic acids are useful for  
 CC identifying proteins or screening for homologous nucleic acids required  
 CC for cellular proliferation to isolate candidate molecules for rational  
 CC drug discovery programs; or for screening homologous nucleic acids  
 CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,  
 CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is encoded by one of  
 CC the target prokaryotic essential genes. Note: The sequence data for this  
 CC patent did not form part of the printed specification, but was obtained  
 CC in electronic format directly from WIPO at  
 CC ftp.wipo.int/pub/published\_pct\_sequences  
 XX  
 XX Sequence 286 AA;

Query Match 2.9%; Score 7; DB 6; Length 286;  
 Best Local Similarity 100.0%; Pred. No. 7.8e+02;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 126 QPVIGIS 132  
 |||||  
 DB 59 QPVIGIS 65

RESULT 71  
 ABG17587  
 ID ABG17587 standard; protein; 296 AA.  
 XX  
 AC ABG17587;  
 XX  
 DT 18-FEB-2002 (first entry)  
 XX  
 DE Novel human diagnostic protein #17578.  
 XX  
 KW Human; chromosome mapping; gene mapping; gene therapy; forensic;  
 KW food supplement; medical imaging; diagnostic; genetic disorder.  
 XX  
 OS *Homo sapiens*.  
 XX  
 PN WO200175067-A2.  
 XX  
 PD 11-OCT-2001.  
 XX  
 PF 30-MAR-2001; 2001WO-US008631.  
 XX  
 PR 31-MAR-2000; 2000US-00540217.  
 PR 23-AUG-2000; 2000US-00649167.  
 XX  
 PA (HYSB-) HYSEQ INC.  
 XX  
 PI Drmanac RT, Liu C, Tang YT;  
 XX  
 DR WPI; 2001-639362/73.  
 DR N-PSDB; AAS81774.

CC New isolated polynucleotide and encoded polypeptides, useful in  
 PT diagnostics, forensics, gene mapping, identification of mutations  
 PT responsible for genetic disorders or other traits and to assess  
 PT biodiversity.  
 XX  
 PS Claim 20; SEQ ID NO 47946; 103pp; English.  
 XX  
 CC The invention relates to isolated polynucleotide (I) and polypeptide (II)  
 CC sequences. (I) is useful as hybridisation probes, polymerase chain  
 CC reaction (PCR) primers, oligomers, and for chromosome and gene mapping,

CC and in recombinant production of (II). The polynucleotides are also used  
CC in diagnostics as expressed sequence tags for identifying expressed  
CC genes. (I) is useful in gene therapy techniques to restore normal  
CC activity of (II) or to treat disease states involving (II). (II) is  
CC useful for generating antibodies against it, detecting or quantitating a  
CC polypeptide in tissue, as molecular weight markers and as a food  
CC supplement. (II) and its binding partners are useful in medical imaging  
CC of sites expressing (II). (I) and (II) are useful for treating disorders  
CC involving aberrant protein expression or biological activity. The  
CC polypeptide and polynucleotide sequences have applications in  
CC diagnostics, forensics, gene mapping, identification of mutations  
CC responsible for genetic disorders or other traits to assess biodiversity  
CC and to produce other types of data and products dependent on DNA and  
CC amino acid sequences. ABG00010-ABG30377 represent novel human diagnostic  
CC amino acid sequences of the invention. Note: The sequence data for this  
CC patent did not appear in the printed specification, but was obtained in  
CC electronic format directly from WIPO at  
CC ftp.wipo.int/pub/published\_pct\_sequences  
XX  
SQ Sequence 296 AA;

Query Match 2.9%; Score 7; DB 4; Length 296;  
Best Local Similarity 100.0%; Pred. No. 8e+02; Mismatches 0; Indels 0; Gaps 0;  
Matches 7; Conservative 0;

QY 84 TSOALPA 90  
Db 9 TSOALPA 15  
|||||

RESULT 72  
AAU36580  
ID AAU36580 standard; protein; 300 AA.

AC AAU36580;  
XX  
DT 14-FEB-2002 (first entry)

XX Staphylococcus aureus cellular proliferation protein #750.  
XX  
XX Antisense; prokaryotic cellular proliferation protein; antibiotic;  
KW antibacterial; drug design.  
XX  
XX Staphylococcus aureus.

OS  
XX WO200170955-A2.  
PN  
XX 27-SEP-2001.

XX 21-MAR-2001; 2001WO-US009180.  
PF  
XX 21-MAR-2000; 2000US-0191078P.  
PR 23-MAY-2000; 2000US-0206848P.  
PR 26-MAY-2000; 2000US-0207727P.  
PR 23-OCT-2000; 2000US-0242578P.  
PR 27-NOV-2000; 2000US-0253625P.  
PR 22-DEC-2000; 2000US-0257931P.  
PR 16-FEB-2001; 2001US-0269308P.

XX (ELIT-) ELITRA PHARM INC.  
FA  
XX Haselbeck R, Ohlson KL, Zyskind JW, Wall D, Trawick JD, Carr GJ;  
PI Yamamoto RT, Xu HH;  
XX WPI; 2001-611495/70.  
DR N-PSDB; AAS54439.

XX New polynucleotides for the identification and development of  
PT antibiotics, comprise sequences of antisense nucleic acids.  
XX  
PS Example 3; SEQ ID NO 12173; 511pp; English.

XX The invention relates to antisense inhibitors of genes essential to

CC prokaryotic cellular proliferation, their use in identifying the genes,  
CC their use in the discovery of novel antibiotics, the essential genes  
CC themselves and the encoded proteins. The prokaryotes used are Escherichia  
CC coli, Staphylococcus aureus, Salmonella typhi, Klebsiella pneumoniae,  
CC Pseudomonas aeruginosa and Enterococcus faecalis. The invention is also  
CC useful for the identification of potential new targets for antibiotic  
CC development. The antisense nucleic acids can also be used to identify  
CC proteins used in proliferation, to express these proteins, and to obtain  
CC antibodies capable of binding to the expressed proteins. The proteins can  
CC be used to screen compounds in rational drug discovery programmes. The  
CC antisense nucleic acid sequence is also useful to screen for homologous  
CC nucleic acids which are required for cell proliferation in a wide variety  
CC of organisms. The present sequence represents an essential prokaryotic  
CC cellular proliferation protein. Note: The sequence data for this patent  
CC did not form part of the printed specification, but was obtained in  
CC electronic format directly from WIPO at  
CC ftp.wipo.int/pub/published\_pct\_sequences  
XX  
SQ Sequence 300 AA;

Query Match 2.9%; Score 7; DB 4; Length 300;  
Best Local Similarity 100.0%; Pred. No. 8.1e-02; Mismatches 0; Indels 0; Gaps 0;  
Matches 7; Conservative 0;

QY 126 QPVIGIS 132  
Db 73 QPVIGIS 79  
|||||

RESULT 73  
ABM71753  
ID ABM71753 standard; protein; 300 AA.

AC ABM71753;  
XX  
DT 20-NOV-2003 (first entry)

XX Staphylococcus aureus protein #993.  
XX  
XX Antibacterial; vaccine; gene therapy; infection; sepsis; diagnosis;  
KW enzymatic assay; antibiotic target.  
XX  
XX Staphylococcus aureus.

OS  
XX WO200294868-A2.  
PN  
XX 28-NOV-2002.

XX 27-MAR-2002; 2002WO-IB002637.  
PF  
XX 27-MAR-2001; 2001GB-00007661.  
PR (CHIR-) CHIRON SPA.

PA Masignani V, Mora M, Scarselli M;  
PI WPI; 2003-120786/11.  
XX N-PSDB; ACF73313.

XX New Staphylococcus aureus protein, useful as a vaccine for treating or  
PT preventing Staphylococcal infection, specifically an infection caused by  
PT S. aureus, e.g. sepsis.

PS Claim 1; SEQ ID NO 1986; 49pp; English.

XX The invention relates to novel genes and encoded proteins from  
CC Staphylococcus aureus. A composition comprising the S. aureus protein, a  
CC nucleic acid encoding the protein, or an antibody to the protein, is  
CC useful as a pharmaceutical, particularly as a vaccine for treating or  
CC preventing infection due to Staphylococcus bacteria, specifically an  
CC infection caused by S. aureus. The composition is particularly useful for  
CC treating or preventing sepsis in a patient. The composition can also be  
CC used for diagnostics. The protein is also used in an assay for enzymatic

CC studies and as a target for antibiotics. This sequence represents one of  
 CC the novel S. aureus proteins of the invention  
 XX  
 SQ Sequence 300 AA;

Query Match 2.9%; Score 7; DB 6; Length 300;  
 Best Local Similarity 100.0%; Pred. No. 8.1e+02;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 126 QPVIGIS 132  
 DB 73 QPVIGIS 79

RESULT 74  
 ID ABB68531  
 XX ABB68531 standard; protein; 302 AA.  
 AC ABB68531;  
 XX  
 DT 26-MAR-2002 (first entry)  
 DE Drosophila melanogaster polypeptide SEQ ID NO 32385.  
 XX Drosophila; developmental biology; cell signalling; insecticide;  
 KW pharmaceutical.  
 XX  
 OS Drosophila melanogaster.  
 XX  
 PN WO200171042-A2.  
 XX  
 PD 27-SEP-2001.  
 XX  
 PF 23-MAR-2001; 2001WO-US009231.  
 XX  
 PR 23-MAR-2000; 2000US-0191637P.  
 XX  
 PR 11-JUL-2000; 2000US-00614150.  
 XX  
 PA (PEKE ) PE CORP NY.  
 XX  
 PI Venter JC, Adams M, Li FWD, Myers EW;  
 XX  
 DR WPI; 2001-656860/75.  
 DR N-PSDB; ABLI2634.  
 XX  
 PT New isolated nucleic acid detection reagent for detecting 1000 or more  
 PT genes from Drosophila and for elucidating cell signalling and cell-cell  
 PT interactions.

XX Disclosure; SEQ ID NO 32385; 21pp + Sequence Listing; English.  
 CC The invention relates to an isolated nucleic acid detection reagent  
 CC capable of detecting 1000 or more genes from Drosophila. The invention is  
 CC useful in developmental biology and in elucidating cell signalling and  
 CC cell-cell interactions in higher eukaryotes for the development of  
 CC insecticides, therapeutics and pharmaceutical drugs. The invention  
 CC discloses genomic DNA sequences (ABLI6176-ABLI30511), expressed DNA  
 CC sequences (ABLI01840-ABLI6175) and the encoded proteins (ABBS57737-  
 CC ABB72072). The sequence data for this patent did not form part of the  
 CC printed specification, but was obtained in electronic format directly  
 CC from WIPO at ftp.wipo.int/pub/published\_pct\_sequences  
 XX  
 SQ Sequence 302 AA;

Query Match 2.9%; Score 7; DB 4; Length 302;  
 Best Local Similarity 100.0%; Pred. No. 8.2e+02;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 64 SGEAGVP 70  
 DB 176 SGEAGVP 182

RESULT 75  
 ID AAY06112  
 XX AAY06112 standard; protein; 323 AA.  
 AC AAY06112;  
 XX  
 DT 16-AUG-1999 (first entry)  
 DE Nelson Bay virus sigma 3 protein.  
 XX  
 KW Orthoreovirus; reovirus; NBV; sigma 3; viral cell attachment;  
 cell fusion.  
 XX  
 OS Nelson Bay virus.  
 XX  
 PN WO9324582-A1.  
 XX  
 PD 20-MAY-1999.  
 XX  
 PF 06-NOV-1998; 98WO-CA001046.  
 XX  
 PR 07-NOV-1997; 97US-00965708.  
 XX  
 PA (UYDA-) UNIV DALHOUSIE.  
 XX  
 PI Duncan R;  
 XX  
 DR WPI; 1999-327410/27.  
 DR N-PSDB; AAX58669.  
 XX  
 PT Fusogenic proteins from non-enveloped reoviruses.  
 XX  
 PS Example 6; Page 51-52; 57pp; English.  
 XX  
 CC The present sequence represents sigma 3 protein, the viral cell  
 CC attachment protein, of Nelson Bay virus (NBV). The coding region for this  
 CC protein lies downstream of the fusogenic P11 protein coding region in the  
 CC NBV genome. 2 Unrelated fusion proteins responsible for cell-cell fusion  
 CC induced by avian reovirus (ARV) and the only 2 fusogenic mammalian  
 CC reoviruses, NBV and baboon reovirus (BRV), are identified in this  
 CC invention. These proteins are named P11 for ARV and NBV (see AAY06104,  
 CC AAY06107 and AAY06110) and P15 for BRV (see AAY06113). Fusogenic P11 and  
 CC P15 proteins are useful for promoting cell and liposome fusion, e.g. for  
 CC production of hybridomas and for liposome-mediated delivery of bioactive  
 CC agents to cells  
 XX  
 SQ Sequence 323 AA;

Query Match 2.9%; Score 7; DB 2; Length 323;  
 Best Local Similarity 100.0%; Pred. No. 8.7e+02;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 31 AQSGLAS 37  
 DB 118 AQSGLAS 124

Search completed: March 17, 2004, 07:24:49  
 Job time : 65 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 18, 2004, 13:17:24 ; Search time 42 Seconds  
(without alignments)  
1492.076 Million cell updates/sec

Title: US-10-057-951-2\_COPY\_22\_263

Perfect score: 242

Sequence: 1 SGCGFWNGHLYREDQTSPL.....PVDQEGSTPLMGQAGTPCA 242

Scoring table: OLIGO

Gapop 60.0 , Gapext 60.0

Searched: 1049977 seqs, 258955339 residues

Word size : 0

Total number of hits satisfying chosen parameters: 1049977

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 100 summaries

Databases : Published Applications AA:  
1: /cgn2\_5/prodata/2/pubaa/US07\_PUBCOMB.pep.\*  
2: /cgn2\_6/prodata/2/pubaa/PCT\_NEW\_PUB.pep.\*  
3: /cgn2\_6/prodata/2/pubaa/US06\_NEW\_PUB.pep.\*  
4: /cgn2\_6/prodata/2/pubaa/US06\_PUBCOMB.pep.\*  
5: /cgn2\_6/prodata/2/pubaa/PCTUS\_NEW\_PUB.pep.\*  
6: /cgn2\_6/prodata/2/pubaa/PCTUS\_PUBCOMB.pep.\*  
7: /cgn2\_6/prodata/2/pubaa/US08\_NEW\_PUB.pep.\*  
8: /cgn2\_6/prodata/2/pubaa/US08\_PUBCOMB.pep.\*  
9: /cgn2\_6/prodata/2/pubaa/US09\_PUBCOMB.pep.\*  
10: /cgn2\_6/prodata/2/pubaa/US09\_PUBCOMB.pep.\*  
11: /cgn2\_6/prodata/2/pubaa/US09\_PUBCOMB.pep.\*  
12: /cgn2\_6/prodata/2/pubaa/US09\_NEW\_PUB.pep.\*  
13: /cgn2\_6/prodata/2/pubaa/US10\_PUBCOMB.pep.\*  
14: /cgn2\_6/prodata/2/pubaa/US10\_PUBCOMB.pep.\*  
15: /cgn2\_6/prodata/2/pubaa/US10\_PUBCOMB.pep.\*  
16: /cgn2\_6/prodata/2/pubaa/US10\_NEW\_PUB.pep.\*  
17: /cgn2\_6/prodata/2/pubaa/US60\_NEW\_PUB.pep.\*  
18: /cgn2\_6/prodata/2/pubaa/US60\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	242	100.0	263	9	US-09-084-491A-2
2	242	100.0	263	13	US-10-102-704-2
3	242	100.0	263	13	US-10-057-951-2
4	229	94.6	263	14	US-10-210-951-44
5	229	94.6	263	14	US-10-211-884-44
6	66	27.3	66	9	US-09-864-761-38457
7	9	3.7	331	12	US-10-282-122A-44397
8	8	3.3	155	12	US-10-425-114-46168
9	8	3.3	457	14	US-10-156-761-10667
10	8	3.3	602	12	US-10-282-122A-51503
11	8	3.3	716	11	US-09-581-151A-55
12	8	3.3	716	12	US-09-972-211-135
13	8	3.3	771	12	US-10-282-122A-62527
14	8	3.3	771	12	US-10-282-122A-64470
15	8	3.3	774	12	US-10-282-122A-62240

16	3.3	3507	15	US-10-369-493-5784	Sequence 5784, Ap
17	2.9	43	9	US-09-864-761-45155	Sequence 45155, A
18	2.9	56	12	US-10-424-599-280850	Sequence 280850, A
19	2.9	69	14	US-10-029-388-27831	Sequence 27831, A
20	2.9	70	12	US-10-424-599-285193	Sequence 285193, A
21	2.9	86	12	US-10-424-599-251292	Sequence 251292, A
22	2.9	106	12	US-10-424-599-193606	Sequence 193606, A
23	2.9	111	14	US-10-279-579A-18	Sequence 18, Appl
24	2.9	133	12	US-10-424-599-160298	Sequence 160298, A
25	2.9	134	12	US-10-424-599-224277	Sequence 224277, A
26	2.9	137	12	US-10-424-599-249048	Sequence 249048, A
27	2.9	141	15	US-10-108-260A-3939	Sequence 3939, Ap
28	2.9	153	12	US-10-282-122A-55102	Sequence 55102, A
29	2.9	154	9	US-09-738-626-3789	Sequence 3789, Ap
30	2.9	182	15	US-10-289-762-846	Sequence 846, App
31	2.9	210	12	US-10-424-599-175867	Sequence 175867, A
32	2.9	228	15	US-10-289-762-849	Sequence 849, App
33	2.9	237	9	US-09-764-868-727	Sequence 727, App
34	2.9	237	11	US-09-764-875-891	Sequence 891, App
35	2.9	237	11	US-09-764-875-1167	Sequence 1167, Ap
36	2.9	258	9	US-09-815-242-5680	Sequence 5680, Ap
37	2.9	273	12	US-10-424-599-284036	Sequence 284036, A
38	2.9	286	12	US-10-282-122A-44430	Sequence 44430, A
39	2.9	300	9	US-09-815-242-12173	Sequence 12173, A
40	2.9	318	12	US-10-425-114-61719	Sequence 61719, A
41	2.9	323	9	US-09-943-002-12	Sequence 12, Appl
42	2.9	330	15	US-10-369-493-356	Sequence 356, App
43	2.9	330	15	US-10-369-493-21323	Sequence 21323, A
44	2.9	335	9	US-09-815-242-11065	Sequence 11065, A
45	2.9	335	12	US-10-282-122A-58236	Sequence 58236, A
46	2.9	338	15	US-10-408-456-22	Sequence 22, Appl
47	2.9	338	15	US-10-408-456-24	Sequence 24, Appl
48	2.9	338	15	US-10-408-456-26	Sequence 26, Appl
49	2.9	355	12	US-10-282-122A-57954	Sequence 57954, A
50	2.9	356	15	US-10-312-273-63	Sequence 63, Appl
51	2.9	365	15	US-10-312-273-81	Sequence 81, Appl
52	2.9	372	12	US-10-425-114-42346	Sequence 42346, A
53	2.9	396	11	US-09-764-875-779	Sequence 779, App
54	2.9	399	9	US-09-764-868-850	Sequence 850, App
55	2.9	419	12	US-10-282-122A-73105	Sequence 73105, A
56	2.9	424	12	US-10-282-122A-48132	Sequence 48132, A
57	2.9	432	15	US-10-369-493-12020	Sequence 12020, A
58	2.9	442	15	US-10-369-493-11707	Sequence 11707, A
59	2.9	442	15	US-10-369-493-12641	Sequence 12641, A
60	2.9	463	12	US-10-282-122A-75534	Sequence 75534, A
61	2.9	465	9	US-09-815-242-13961	Sequence 13961, A
62	2.9	465	12	US-10-282-122A-55975	Sequence 55975, A
63	2.9	489	12	US-10-282-122A-43395	Sequence 43395, A
64	2.9	469	12	US-10-282-122A-50382	Sequence 50382, A
65	2.9	501	15	US-10-408-456-28	Sequence 28, Appl
66	2.9	545	14	US-10-156-761-11459	Sequence 11459, A
67	2.9	552	12	US-10-425-114-67035	Sequence 67035, A
68	2.9	562	9	US-09-738-626-5317	Sequence 5317, Ap
69	2.9	563	12	US-10-424-599-146903	Sequence 146903, A
70	2.9	575	15	US-10-369-493-19493	Sequence 19493, A
71	2.9	684	9	US-09-946-678-2	Sequence 2, Appl
72	2.9	701	9	US-09-815-242-13411	Sequence 13411, A
73	2.9	701	12	US-10-282-122A-73782	Sequence 73782, A
74	2.9	706	9	US-09-738-626-6019	Sequence 6019, Ap
75	2.9	774	12	US-10-425-114-43637	Sequence 43637, A
76	2.9	779	12	US-10-282-122A-69134	Sequence 69134, A
77	2.9	822	12	US-10-425-114-62784	Sequence 62784, A
78	2.9	885	15	US-10-369-493-20661	Sequence 20661, Ap
79	2.9	937	9	US-09-974-298-129	Sequence 129, App
80	2.9	1113	12	US-10-282-122A-66977	Sequence 66977, A
81	2.9	1156	14	US-10-099-285-72	Sequence 72, Appl
82	2.9	1156	15	US-10-428-961-28	Sequence 28, Appl
83	2.9	1286	15	US-10-369-493-5431	Sequence 5431, Ap
84	2.9	2472	9	US-09-815-242-5064	Sequence 5064, Ap
85	2.5	7	10	US-09-792-286-183	Sequence 183, App
86	2.5	7	10	US-09-792-286-183	Sequence 258, App
87	2.5	7	10	US-09-792-286-183	Sequence 74, Appl
88	2.5	16	14	US-10-206-699-74	

89 6 2.5 20 14 US-10-032-221B-28 Sequence 28, Appl  
90 6 2.5 26 12 US-10-306-631-91 Sequence 91, Appl  
91 6 2.5 27 9 US-09-864-761-35595 Sequence 35595, A  
92 6 2.5 28 12 US-10-306-631-92 Sequence 92, Appl  
93 6 2.5 32 9 US-09-864-761-48415 Sequence 48415, A  
94 6 2.5 34 10 US-09-820-649-121 Sequence 121, App  
95 6 2.5 34 14 US-10-160-162-121 Sequence 121, App  
96 6 2.5 35 12 US-10-424-599-248913 Sequence 248913, App  
97 6 2.5 37 14 US-10-106-698-8286 Sequence 8286, App  
98 6 2.5 38 9 US-09-864-761-38182 Sequence 38182, A  
99 6 2.5 38 12 US-10-306-631-93 Sequence 93, Appl  
100 6 2.5 39 10 US-09-983-802-589 Sequence 589, App

# ALIGNMENTS

RESULT 1  
US-09-084-491A-2  
; Sequence 2, Application US/09084491A  
; Patent No. US20020061576A1  
; GENERAL INFORMATION:  
; APPLICANT: MOORE, PAUL A.  
; APPLICANT: RUBEN, STEVEN M.  
; APPLICANT: RUBEN, REINHARD  
; TITLE OF INVENTION: TISSUE PLASMINOGEN ACTIVATOR-LIKE PROTEASE  
; NUMBER OF SEQUENCES: 16  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: HUMAN GENOME SCIENCES, INC.  
; STREET: 9410 KEY WEST AVENUE  
; CITY: ROCKVILLE  
; STATE: MD  
; COUNTRY: USA  
; ZIP: 20850  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent in Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/084,491A  
; FILING DATE: 27-MAY-1998  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: BROOKES, ANDERS A.  
; REGISTRATION NUMBER: 36,373  
; REFERENCE/DOCKET NUMBER: PF378  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (301) 309-8504  
; TELEFAX: (301) 309-8439  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 263 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein

US-09-084-491A-2  
Query Match 100.0%; Score 242; DB 9; Length 263;  
Best Local Similarity 100.0%; Pred. No. 2.3e-222; Indels 0; Gaps 0;  
Matches 242; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 SGGCFWNGHLYREDQTSAPGLRCLNWLDAQSGLASAPVSGAGNHSYCRNPDEDPGFW 60  
DB 22 SGGCFWNGHLYREDQTSAPGLRCLNWLDAQSGLASAPVSGAGNHSYCRNPDEDPGFW 81  
QY 61 CYVSGAGVPEKPCEDLRCPETTSQALPAFTTEQAESEGGCADEVQVAFANALPARS 120  
DB 82 CYVSGAGVPEKPCEDLRCPETTSQALPAFTTEQAESEGGCADEVQVAFANALPARS 141  
QY 121 EAAAVQPVIGISQVRMNSKEKDLGLTGLVGLITMWVIIAIGAGIILGYSYKRGKDLK 180  
DB 142 EAAAVQPVIGISQVRMNSKEKDLGLTGLVGLITMWVIIAIGAGIILGYSYKRGKDLK 201

QY 181 EQHDKVCEREMORITLPLSAFTNPTCEIVDEKTVVHTSQTVPDQSGSTPLMGQAGTP 240  
DB 202 EQHDKVCEREMORITLPLSAFTNPTCEIVDEKTVVHTSQTVPDQSGSTPLMGQAGTP 261  
QY 241 GA 242  
DB 262 GA 263

RESULT 2  
US-10-102-704-2  
; Sequence 2, Application US/10102704  
; Publication No. US20020164768A1  
; GENERAL INFORMATION:  
; APPLICANT: Moore et al.  
; TITLE OF INVENTION: Tissue Plasminogen Activator-Like Protein  
; FILE REFERENCE: PF378C1  
; CURRENT APPLICATION NUMBER: US/10/102,704  
; PRIOR FILING DATE: 2002-03-22  
; PRIOR APPLICATION NUMBER: 09/084,491  
; PRIOR FILING DATE: 1998-05-27  
; PRIOR APPLICATION NUMBER: 60/048,000  
; PRIOR FILING DATE: 1997-05-28  
; NUMBER OF SEQ ID NOS: 16  
; SOFTWARE: Patent in version 3.1  
; SEQ ID NO 2  
; LENGTH: 263  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-102-704-2

Query Match 100.0%; Score 242; DB 13; Length 263;  
Best Local Similarity 100.0%; Pred. No. 2.3e-222; Indels 0; Gaps 0;  
Matches 242; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SGGCFWNGHLYREDQTSAPGLRCLNWLDAQSGLASAPVSGAGNHSYCRNPDEDPGFW 60  
DB 22 SGGCFWNGHLYREDQTSAPGLRCLNWLDAQSGLASAPVSGAGNHSYCRNPDEDPGFW 81  
QY 61 CYVSGAGVPEKPCEDLRCPETTSQALPAFTTEQAESEGGCADEVQVAFANALPARS 120  
DB 82 CYVSGAGVPEKPCEDLRCPETTSQALPAFTTEQAESEGGCADEVQVAFANALPARS 141  
QY 121 EAAAVQPVIGISQVRMNSKEKDLGLTGLVGLITMWVIIAIGAGIILGYSYKRGKDLK 180  
DB 142 EAAAVQPVIGISQVRMNSKEKDLGLTGLVGLITMWVIIAIGAGIILGYSYKRGKDLK 201  
QY 181 EQHDKVCEREMORITLPLSAFTNPTCEIVDEKTVVHTSQTVPDQSGSTPLMGQAGTP 240  
DB 202 EQHDKVCEREMORITLPLSAFTNPTCEIVDEKTVVHTSQTVPDQSGSTPLMGQAGTP 261  
QY 241 GA 242  
DB 262 GA 263

RESULT 3  
US-10-057-951-2  
; Sequence 2, Application US/10057951  
; Publication No. US20020177213A1  
; GENERAL INFORMATION:  
; APPLICANT: Moore et al.  
; TITLE OF INVENTION: Tissue Plasminogen Activator-Like Protease  
; FILE REFERENCE: PF378P1  
; CURRENT APPLICATION NUMBER: US/10/057,951  
; CURRENT FILING DATE: 2002-01-29  
; PRIOR FILING DATE: 2002-01-29  
; PRIOR APPLICATION NUMBER: US 09/411,977  
; PRIOR FILING DATE: 1999-10-04  
; PRIOR APPLICATION NUMBER: US 09/084,491  
; PRIOR FILING DATE: 1998-05-27  
; PRIOR APPLICATION NUMBER: US 60/048,000  
; PRIOR FILING DATE: 1997-05-28

```
; NUMBER OF SEQ ID NOS: 31
; SOFTWARE: Patentin Ver. 3.1
; SEQ ID NO 2
; LENGTH: 263
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-057-951-2

Query Match      100.0%; Score 242; DB 13; Length 263;
Best Local Similarity 100.0%; Pred. No. 2,3e-222;
Matches 242; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SGGCFWNGHLYREDQTSAPAGLRCLNWLDAQSLASAPVSGAGNHSYCRNPDEDPGFW 60
Db 22 SGGCFWNGHLYREDQTSAPAGLRCLNWLDAQSLASAPVSGAGNHSYCRNPDEDPGFW 81
QY 61 CYSGEAGVPEKPCEDLRCPETTSOALPAFTTEIQEASEGPGADDEVQVFAPANALPARS 120
Db 82 CYSGEAGVPEKPCEDLRCPETTSOALPAFTTEIQEASEGPGADDEVQVFAPANALPARS 141
QY 121 EAAAVQPVIGISQVRVMSKXKDLGLTGLGYVLGITVMVIIAIGAGIILGYSYKRGKDLK 180
Db 142 EAAAVQPVIGISQVRVMSKXKDLGLTGLGYVLGITVMVIIAIGAGIILGYSYKRGKDLK 201
QY 181 EQHDKVCEREMQRIITLPLSAFTNPTCEIVDEKTVVHTSOTPVDPQEGSTPLMGAGTP 240
Db 202 EQHDKVCEREMQRIITLPLSAFTNPTCEIVDEKTVVHTSOTPVDPQEGSTPLMGAGTP 261
QY 241 GA 242
Db 262 GA 263

RESULT 4
US-10-210-951-44
; Sequence 44, Application US/10210951
; Publication No. US20030170228A1
; GENERAL INFORMATION:
; APPLICANT: Ashkenazi, Avi J.
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Hillan, Kenneth J.
; APPLICANT: Marsters, Scot A.
; APPLICANT: Pan, James
; APPLICANT: Pitti, Robert M.
; APPLICANT: Roy, Margaret Ann
; APPLICANT: Smith, Victoria
; APPLICANT: Stone, Donna M.
; APPLICANT: Watanabe, Colin K.
; APPLICANT: Wood, William I.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE TREATMENT OF TUMOR
; FILE REFERENCE: P2931R1C1
; CURRENT APPLICATION NUMBER: US/10/210,951
; CURRENT FILING DATE: 2002-08-02
; PRIOR APPLICATION NUMBER: 60/014699
; PRIOR FILING DATE: 1996-04-01
; PRIOR APPLICATION NUMBER: 60/026943
; PRIOR FILING DATE: 1996-09-23
; PRIOR APPLICATION NUMBER: 60/059121
; PRIOR FILING DATE: 1997-07-17
; PRIOR APPLICATION NUMBER: 60/059352
; PRIOR FILING DATE: 1997-09-19
; PRIOR APPLICATION NUMBER: 60/062037
; PRIOR FILING DATE: 1997-10-10
; PRIOR APPLICATION NUMBER: 60/063755
; PRIOR FILING DATE: 1997-10-10
; PRIOR APPLICATION NUMBER: 60/063755
; PRIOR FILING DATE: 1997-10-17
; PRIOR APPLICATION NUMBER: 60/063045
; PRIOR FILING DATE: 1997-10-17
; PRIOR APPLICATION NUMBER: 60/063046
; PRIOR FILING DATE: 1997-10-24
; PRIOR APPLICATION NUMBER: 60/063046
; PRIOR FILING DATE: 1997-10-24
; PRIOR APPLICATION NUMBER: 60/066511
; PRIOR FILING DATE: 1997-10-24
; PRIOR APPLICATION NUMBER: 60/066511
; PRIOR FILING DATE: 1997-11-24

US-10-057-951-44
Query Match      94.6%; Score 229; DB 14; Length 263;
Best Local Similarity 100.0%; Pred. No. 5.8e-210;
Matches 229; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SGGCFWNGHLYREDQTSAPAGLRCLNWLDAQSLASAPVSGAGNHSYCRNPDEDPGFW 60
Db 22 SGGCFWNGHLYREDQTSAPAGLRCLNWLDAQSLASAPVSGAGNHSYCRNPDEDPGFW 81
QY 61 CYSGEAGVPEKPCEDLRCPETTSOALPAFTTEIQEASEGPGADDEVQVFAPANALPARS 120
Db 82 CYSGEAGVPEKPCEDLRCPETTSOALPAFTTEIQEASEGPGADDEVQVFAPANALPARS 141
QY 121 EAAAVQPVIGISQVRVMSKXKDLGLTGLGYVLGITVMVIIAIGAGIILGYSYKRGKDLK 180
Db 142 EAAAVQPVIGISQVRVMSKXKDLGLTGLGYVLGITVMVIIAIGAGIILGYSYKRGKDLK 201
QY 181 EQHDKVCEREMQRIITLPLSAFTNPTCEIVDEKTVVHTSOTPVDPQEG 229
Db 202 EQHDKVCEREMQRIITLPLSAFTNPTCEIVDEKTVVHTSOTPVDPQEG 250

RESULT 5
US-10-211-884-44
; Sequence 44, Application US/10211884
; Publication No. US20030175900A1
; GENERAL INFORMATION:
; APPLICANT: Ashkenazi, Avi J.
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Hillan, Kenneth J.
; APPLICANT: Marsters, Scot A.
; APPLICANT: Pan, James
; APPLICANT: Pitti, Robert M.
; APPLICANT: Roy, Margaret Ann
; APPLICANT: Smith, Victoria
; APPLICANT: Stone, Donna M.
; APPLICANT: Watanabe, Colin K.
; APPLICANT: Wood, William I.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE TREATMENT OF TUMOR
; FILE REFERENCE: P2931R1C1
; CURRENT APPLICATION NUMBER: US/10/211,884
; CURRENT FILING DATE: 2002-08-02
; PRIOR APPLICATION NUMBER: 60/014699
; PRIOR FILING DATE: 1996-04-01
; PRIOR APPLICATION NUMBER: 60/026943
; PRIOR FILING DATE: 1996-09-23
; PRIOR APPLICATION NUMBER: 60/059121
; PRIOR FILING DATE: 1997-07-17
; PRIOR APPLICATION NUMBER: 60/059352
; PRIOR FILING DATE: 1997-09-19
; PRIOR APPLICATION NUMBER: 60/062037
; PRIOR FILING DATE: 1997-10-10
; PRIOR APPLICATION NUMBER: 60/063755
; PRIOR FILING DATE: 1997-10-17
; PRIOR APPLICATION NUMBER: 60/063045
; PRIOR FILING DATE: 1997-10-24
; PRIOR APPLICATION NUMBER: 60/063046
; PRIOR FILING DATE: 1997-10-24
; PRIOR APPLICATION NUMBER: 60/066511
; PRIOR FILING DATE: 1997-11-24
; PRIOR APPLICATION NUMBER: 60/066772
; PRIOR FILING DATE: 1997-11-24
```

```

; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 258
; SEQ ID NO 44
; LENGTH: 263
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-211-884-44

Query Match
Best Local Similarity 100.0%; Pred. No. 5.8e-210; Indels 0; Gaps 0;
Matches 229; Conservative 0; Mismatches 0;

QY 1 SGGCFWNGHLYREDQTSAPGLRCLNWLDAQGLASAPVSGAGNHSYCRNPDEDPRGPW 60
Db 22 SGGCFWNGHLYREDQTSAPGLRCLNWLDAQGLASAPVSGAGNHSYCRNPDEDPRGPW 81

QY 61 CYSVGBAGVPEKRPEDLRCPTTSQALPAFTTEIQEASEGPGDEVOVFAPANALPARS 120
Db 82 CYSVGBAGVPEKRPEDLRCPTTSQALPAFTTEIQEASEGPGDEVOVFAPANALPARS 141

QY 121 EAAAVOPVIGISQVRMNSKEKDLGTLGVILGIMVWIIAIGAGIILGYSYKRGKDLK 180
Db 142 EAAAVOPVIGISQVRMNSKEKDLGTLGVILGIMVWIIAIGAGIILGYSYKRGKDLK 201

QY 181 EHQDKVCEREMQRIITPLSAFTNPTCEIVDEKTVVWHTSQTPVDPQEG 229
Db 202 EHQDKVCEREMQRIITPLSAFTNPTCEIVDEKTVVWHTSQTPVDPQEG 250

RESULT 6
US-09-864-761-38457
; Sequence 38457, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharon G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY
; FILE REFERENCE: Aeonica-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30

```

```

; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Annonax Sequence Listing Engine vers. 1.1
; SEQ ID NO 38457
; LENGTH: 66
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AC02073.1
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 2.2
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 2.5
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 2.3
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 1.9
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 2.2
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 2.2
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 2.6
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 2.2
; OTHER INFORMATION: SWISSPROT HIT: P39881, EVALUE 3.40e+00
; OTHER INFORMATION: EST_HUMAN HIT: BE299406.1, EVALUE 5.00e-30
US-09-864-761-38457

Query Match
Best Local Similarity 100.0%; Pred. No. 4.8e-55;
Matches 66; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 83 TTSQLPAFTTEIQEASEGPGDEVOVFAPANALPARSAAAVOPVIGISQVRMNSKEK 142
Db 1 TTSQLPAFTTEIQEASEGPGDEVOVFAPANALPARSAAAVOPVIGISQVRMNSKEK 60

QY 143 KDLGTL 148
Db 61 KDLGTL 66

RESULT 7
US-10-282-122A-44397
; Sequence 44397, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zvekind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625

```



```
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 44397
; LENGTH: 331
; TYPE: PRT
; ORGANISM: Staphylococcus aureus
US-10-282-122A-44397

Query Match          3.7%; Score 9; DB 12; Length 331;
Best Local Similarity 100.0%; Pred.No. 4.5;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 160 IIAIGAGII 168
Db 97 IIAIGAGII 105

RESULT 8
US-10-425-114-46168
; Sequence 46168, Application US/10425114
; Publication No. US20040034888A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Jingdong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E.
; APPLICANT: Tabaska, Jack E
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53313)B
; CURRENT APPLICATION NUMBER: US/10/425,114
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 46168
; LENGTH: 155
; TYPE: PRT
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: 701165192_FLI.pep
US-10-425-114-46168

Query Match          3.3%; Score 8; DB 12; Length 155;
Best Local Similarity 100.0%; Pred.No. 21;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 37 SAPVSGAG 44
Db 15 SAPVSGAG 22

RESULT 9
US-10-156-761-10667
; Sequence 10667, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
```

```
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 10667
; LENGTH: 457
; TYPE: PRT
; ORGANISM: Streptomyces avermitilis
US-10-156-761-10667

Query Match          3.3%; Score 8; DB 14; Length 457;
Best Local Similarity 100.0%; Pred.No. 54;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 63 VSGEAGVP 70
Db 352 VSGEAGVP 359

RESULT 10
US-10-282-122A-51503
; Sequence 51503, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 51503
; LENGTH: 602
; TYPE: PRT
; ORGANISM: Clostridium acetobutylicum
US-10-282-122A-51503

Query Match          3.3%; Score 8; DB 12; Length 602;
```

Best Local Similarity 100.0%; Pred. No. 69;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 63 VSGBAGVP 70  
|||||  
Db 213 VSGBAGVP 220

RESULT 11  
US-09-981-151A-55  
; Sequence 55, Application US/09981151A  
; Publication No. US2003021256A1  
; GENERAL INFORMATION:  
; APPLICANT: Edinger, Shlomit R  
; APPLICANT: Gerlach, Valerie  
; APPLICANT: MacDougall, John R  
; APPLICANT: Malvankar, Muriel M  
; APPLICANT: Smithson, Glennda  
; APPLICANT: Millet, Isabelle  
; APPLICANT: Peyman, John A  
; APPLICANT: Stone, David J  
; APPLICANT: Gunther, Erik  
; APPLICANT: Ellerman, Karen  
; APPLICANT: Shimkets, Richard A  
; APPLICANT: Padigaru, Muralidhara  
; APPLICANT: Guo, Xiaojia  
; APPLICANT: Patturajan, Meera  
; APPLICANT: Taupier Jr, Raymond J  
; APPLICANT: Burgess, Catherine E  
; APPLICANT: Zerhusen, Bryan D  
; APPLICANT: Kekuda, Ramesh  
; APPLICANT: Spytek, Kimberly A  
; APPLICANT: Gangolli, Bha A  
; APPLICANT: Fernandes, Elma R  
; APPLICANT: Gorman, Linda

; TITLE OF INVENTION: Proteins and Nucleic Acids Encoding Same  
; FILE REFERENCE: 21402-168  
; CURRENT FILING DATE: 2001-10-16  
; PRIOR APPLICATION NUMBER: US/09/981,151A  
; CURRENT FILING DATE: 2001-10-16  
; PRIOR APPLICATION NUMBER: 60/241,040  
; PRIOR FILING DATE: 2000-10-17  
; PRIOR APPLICATION NUMBER: 60/241,058  
; PRIOR FILING DATE: 2000-10-17  
; PRIOR APPLICATION NUMBER: 60/241,063  
; PRIOR FILING DATE: 2000-10-17  
; PRIOR APPLICATION NUMBER: 60/241,243  
; PRIOR FILING DATE: 2000-10-17  
; PRIOR APPLICATION NUMBER: 60/242,152  
; PRIOR FILING DATE: 2000-10-20  
; PRIOR APPLICATION NUMBER: 60/242,482  
; PRIOR FILING DATE: 2000-10-23  
; PRIOR APPLICATION NUMBER: 60/242,611  
; PRIOR FILING DATE: 2000-10-23  
; PRIOR APPLICATION NUMBER: 60/242,612  
; PRIOR FILING DATE: 2000-10-23  
; PRIOR APPLICATION NUMBER: 60/242,880  
; PRIOR FILING DATE: 2000-10-24  
; PRIOR APPLICATION NUMBER: 60/242,881  
; PRIOR FILING DATE: 2000-10-24  
; Remaining Prior Application data removed - See File Wrapper or PALM.  
; NUMBER OF SEQ ID NOS: 160  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 55  
; LENGTH: 716  
; TYPE: PRT  
; ORGANISM: Mus musculus  
US-09-981-151A-55

Query Match 3.3%; Score 8; DB 11; Length 716;  
Best Local Similarity 100.0%; Pred. No. 81;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 55 DPRGPWCY 62

Db 163 DPRGPWCY 170  
|||||

RESULT 12  
US-09-972-211-135  
; Sequence 135, Application US/09972211  
; Publication No. US20040048245A1  
; GENERAL INFORMATION:  
; APPLICANT: Shimkets, Richard A  
; APPLICANT: Taupier Jr, Raymond J  
; APPLICANT: Burgess, Catherine E  
; APPLICANT: Zerhusen, Bryan D  
; APPLICANT: Mezes, Peter S  
; APPLICANT: Rastelli, Luca  
; APPLICANT: Malvankar, Uriel M  
; APPLICANT: Grosse, William M  
; APPLICANT: Alsbrook II, John P  
; APPLICANT: Lepley, Denise M  
; APPLICANT: Spytek, Kimberly Ann  
; APPLICANT: Li, Li  
; APPLICANT: Edinger, Shlomit  
; APPLICANT: Gerlach, Valerie  
; APPLICANT: Ellerman, Karen  
; APPLICANT: MacDougall, John R  
; APPLICANT: Gunther, Erik  
; APPLICANT: Millet, Isabelle  
; APPLICANT: Stone, David J  
; APPLICANT: Smithson, Glennda  
; APPLICANT: Szekeres Jr, Edward S

; TITLE OF INVENTION: No. US20040048245A1 Human Proteins, Polynucleotides Encoding T  
; FILE REFERENCE: 21402-141  
; CURRENT APPLICATION NUMBER: US/09/972,211  
; CURRENT FILING DATE: 2001-10-05  
; PRIOR APPLICATION NUMBER: 60/238,325  
; PRIOR FILING DATE: 2000-10-05  
; PRIOR APPLICATION NUMBER: 60/238,323  
; PRIOR FILING DATE: 2000-10-05  
; PRIOR APPLICATION NUMBER: 60/238,400  
; PRIOR FILING DATE: 2000-10-06  
; PRIOR APPLICATION NUMBER: 60/238,397  
; PRIOR FILING DATE: 2000-10-06  
; PRIOR APPLICATION NUMBER: 60/238,401  
; PRIOR FILING DATE: 2000-10-06  
; PRIOR APPLICATION NUMBER: 60/238,379  
; PRIOR FILING DATE: 2000-10-06  
; PRIOR APPLICATION NUMBER: 60/238,402  
; PRIOR FILING DATE: 2000-10-06  
; PRIOR APPLICATION NUMBER: 30/238,384  
; PRIOR FILING DATE: 2000-10-06  
; PRIOR APPLICATION NUMBER: 60/238,373  
; PRIOR FILING DATE: 2000-10-06  
; PRIOR APPLICATION NUMBER: 60/238,372  
; PRIOR FILING DATE: 2000-10-06  
; PRIOR APPLICATION NUMBER: 60/238,383  
; PRIOR FILING DATE: 2000-10-06  
; PRIOR APPLICATION NUMBER: 60/238,382  
; PRIOR FILING DATE: 2000-10-06  
; PRIOR APPLICATION NUMBER: 60/275,892  
; PRIOR FILING DATE: 2001-03-14  
; PRIOR APPLICATION NUMBER: 60/236,860  
; PRIOR FILING DATE: 2001-06-08  
; NUMBER OF SEQ ID NOS: 198  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 135  
; LENGTH: 716  
; TYPE: PRT  
; ORGANISM: Mus musculus  
US-09-972-211-135

Query Match 3.3%; Score 8; DB 12; Length 716;  
Best Local Similarity 100.0%; Pred. No. 81;

```
Matches      8; Conservative      0; Mismatches      0; Indels      0; Gaps      0;

Qy      55 DPGPWCY 62
      |||||
Db      163 DPGPWCY 170

RESULT 13
US-10-282-122A-62527
; Sequence 62527, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: Patent in version 3.1
; SEQ ID NO 62527
; LENGTH: 771
; TYPE: PRT
; ORGANISM: Mycobacterium bovis
US-10-282-122A-62527

Query Match      3.3%; Score 8; DB 12; Length 771;
Best Local Similarity 100.0%; Pred. No. 86;
Matches      8; Conservative      0; Mismatches      0; Indels      0; Gaps      0;

Qy      37 SAPVSGAG 44
      |||||
Db      692 SAPVSGAG 699

US-10-282-122A-62527
; Sequence 62527, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20

RESULT 14
US-10-282-122A-64470
; Sequence 64470, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
```

PRIOR APPLICATION NUMBER: 60/191,078  
PRIOR FILING DATE: 2000-03-21  
PRIOR APPLICATION NUMBER: 60/206,848  
PRIOR FILING DATE: 2000-05-23  
PRIOR APPLICATION NUMBER: 60/207,727  
PRIOR FILING DATE: 2000-05-26  
PRIOR APPLICATION NUMBER: 60/230,335  
PRIOR FILING DATE: 2000-09-06  
PRIOR APPLICATION NUMBER: 60/230,347  
PRIOR FILING DATE: 2000-09-09  
PRIOR APPLICATION NUMBER: 60/242,578  
PRIOR FILING DATE: 2000-10-23  
PRIOR APPLICATION NUMBER: 60/253,625  
PRIOR FILING DATE: 2000-11-27  
PRIOR APPLICATION NUMBER: 60/257,931  
PRIOR FILING DATE: 2000-12-22  
PRIOR APPLICATION NUMBER: 60/267,636  
PRIOR FILING DATE: 2001-02-09  
PRIOR APPLICATION NUMBER: 60/269,308  
PRIOR FILING DATE: 2001-02-16  
Remaining Prior Application data removed - See File Wrapper or PALM.  
NUMBER OF SEQ ID NOS: 78614  
SOFTWARE: Patent in version 3.1  
SEQ ID NO 62240  
LENGTH: 774  
TYPE: PRT  
ORGANISM: Mycobacterium avium  
US-10-282-122A-62240

Query Match 3.3%; Score 8; DB 12; Length 774;  
Best Local Similarity 100.0%; Pred. No. 86;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 37 SAPVSGAG 44  
|||||  
DB 695 SAPVSGAG 702

RESULT 16  
US-10-369-493-5784  
Sequence 5784, Application US/10369493  
Publication No. US20030233675A1  
GENERAL INFORMATION:  
APPLICANT: Cao, Yongwei  
APPLICANT: Hinkle, Gregory J.  
APPLICANT: Slater, Steven C.  
APPLICANT: Goldman, Barry S.  
APPLICANT: Chen, Xianfeng  
TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF  
FILE REFERENCE: 38-10(52052)B  
CURRENT APPLICATION NUMBER: US/10/369,493  
CURRENT FILING DATE: 2003-02-28  
PRIOR APPLICATION NUMBER: US 60/360,039  
PRIOR FILING DATE: 2002-02-21  
NUMBER OF SEQ ID NOS: 47374  
SEQ ID NO 5784  
LENGTH: 3507  
TYPE: PRT  
ORGANISM: Caenorhabditis elegans  
US-10-369-493-5784

Query Match 3.3%; Score 8; DB 15; Length 3507;  
Best Local Similarity 100.0%; Pred. No. 3.3e+02;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 64 SGEAGVPE 71  
|||||  
DB 1269 SGEAGVPE 1276

RESULT 17  
US-09-864-761-45155

Sequence 45155, Application US/09864761  
Patent No. US20020048763A1  
GENERAL INFORMATION:  
APPLICANT: Penn, Sharron G.  
APPLICANT: Rank, David R.  
APPLICANT: Hanzel, David K.  
APPLICANT: Chen, Wensheng  
TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR  
FILE REFERENCE: Aecm1ca-X-1  
CURRENT APPLICATION NUMBER: US/09/864,761  
CURRENT FILING DATE: 2001-05-23  
PRIOR APPLICATION NUMBER: US 60/180,312  
PRIOR FILING DATE: 2000-02-04  
PRIOR APPLICATION NUMBER: US 60/207,456  
PRIOR FILING DATE: 2000-05-26  
PRIOR APPLICATION NUMBER: US 09/632,366  
PRIOR FILING DATE: 2000-08-03  
PRIOR APPLICATION NUMBER: GB 24263.6  
PRIOR FILING DATE: 2000-10-04  
PRIOR APPLICATION NUMBER: US 60/236,359  
PRIOR FILING DATE: 2000-09-27  
PRIOR APPLICATION NUMBER: PCT/US01/00666  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00667  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00664  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00669  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00665  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00668  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00663  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00662  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00661  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00670  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: US 60/234,687  
PRIOR FILING DATE: 2000-09-21  
PRIOR APPLICATION NUMBER: US 09/608,408  
PRIOR FILING DATE: 2000-06-30  
PRIOR APPLICATION NUMBER: US 09/774,203  
PRIOR FILING DATE: 2001-01-29  
NUMBER OF SEQ ID NOS: 49117  
SOFTWARE: Annonax Sequence Listing Engine vers. 1.1  
SEQ ID NO 45155  
LENGTH: 43  
TYPE: PRT  
ORGANISM: Homo sapiens  
FEATURE:  
OTHER INFORMATION: MAP TO AC020610.6  
OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 0.75  
OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 0.74  
OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 0.82  
OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 0.96  
OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 0.92  
OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 0.85  
US-09-864-761-45155

Query Match 2.9%; Score 7; DB 9; Length 43;  
Best Local Similarity 100.0%; Pred. No. 60;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 155 TMMVIII 161  
|||||  
DB 26 TMMVIII 32

## RESULT 18

US-10-424-599-280850  
; Sequence 280850, Application US/10424599  
; Publication No. US20040031072A1  
; GENERAL INFORMATION:

APPLICANT: La Rosa Thomas J  
APPLICANT: Kovalic David K  
APPLICANT: Zhou Yihua

TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With  
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement

FILE REFERENCE: 38-21(53223)B  
CURRENT APPLICATION NUMBER: US/10/424,599

CURRENT FILING DATE: 2003-04-28

NUMBER OF SEQ ID NOS: 285684

SEQ ID NO 280850

LENGTH: 56

TYPE: PRT

ORGANISM: Glycine max

FEATURE:

OTHER INFORMATION: Clone ID: PAT\_MRT3847\_9562C.1.pep  
US-10-424-599-280850

Query Match 2.9%; Score 7; DB 12; Length 56;

Best Local Similarity 100.0%; Pred. No. 76;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 212 EKTWVWH 218

Db 4 EKTWVWH 10

## RESULT 19

US-10-029-386-27831  
; Sequence 27831, Application US/10029386  
; Publication No. US20030194704A1  
; GENERAL INFORMATION:

APPLICANT: Penn, Sharon G.  
APPLICANT: Rank, David R.

APPLICANT: Hanzel, David K.

TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR

FILE REFERENCE: AECOMICA-X-2

CURRENT APPLICATION NUMBER: US/10/029,386

CURRENT FILING DATE: 2001-12-20

NUMBER OF SEQ ID NOS: 34288

SOFTWARE: Annonax Sequence Listing Engine vers. 1.1

SEQ ID NO 27831

LENGTH: 69

TYPE: PRT

ORGANISM: Homo sapiens

FEATURE:

OTHER INFORMATION: MAP TO CHR21 66.0

OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1.3

OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 0.89

OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 0.83

OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1

OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.4

OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 2.9

OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.3

OTHER INFORMATION: SWISSPROT HIT: P57075, EVALUE 1.00e-36

US-10-029-386-27831

Query Match 2.9%; Score 7; DB 14; Length 69;

Best Local Similarity 100.0%; Pred. No. 91;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 128 VIGISQR 134

Db 39 VIGISQR 45

## RESULT 20

US-10-424-599-285193  
; Sequence 285193, Application US/10424599  
; Publication No. US20040031072A1  
; GENERAL INFORMATION:

APPLICANT: La Rosa Thomas J  
APPLICANT: Kovalic David K  
APPLICANT: Zhou Yihua

TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With  
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement

FILE REFERENCE: 38-21(53223)B  
CURRENT APPLICATION NUMBER: US/10/424,599

CURRENT FILING DATE: 2003-04-28

NUMBER OF SEQ ID NOS: 285684

SEQ ID NO 285193

LENGTH: 70

TYPE: PRT

ORGANISM: Glycine max

FEATURE:

OTHER INFORMATION: Clone ID: PAT\_MRT3847\_99559C.1.pep  
US-10-424-599-285193

Query Match 2.9%; Score 7; DB 12; Length 86;

Best Local Similarity 100.0%; Pred. No. 1.1e+02;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 113 ANALPAR 119

Db 80 ANALPAR 86

## RESULT 22

US-10-424-599-193606

```

; Sequence 193606, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 193606
; LENGTH: 106
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_16850C.1.pap
US-10-424-599-193606

Query Match      2.9%; Score 7; DB 12; Length 106;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 171 YSYKRGK 177
Db 21 YSYKRGK 27

RESULT 23
US-10-279-579A-18
; Sequence 18, Application US/10279579A
; Publication No. US20030177532A1
; GENERAL INFORMATION:
; APPLICANT: Burrell, Michael
; APPLICANT: Coates, Andrew
; TITLE OF INVENTION: Modification of starch granule size and number
; FILE REFERENCE: 9341-029
; CURRENT APPLICATION NUMBER: US/10/279,579A
; CURRENT FILING DATE: 2002-10-24
; PRIOR APPLICATION NUMBER: 60/346,905
; PRIOR FILING DATE: 2002-01-08
; PRIOR APPLICATION NUMBER: UK 0125493.7
; PRIOR FILING DATE: 2001-10-24
; NUMBER OF SEQ ID NOS: 47
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 18
; LENGTH: 111
; TYPE: PRT
; ORGANISM: Zea mays
US-10-279-579A-18

Query Match      2.9%; Score 7; DB 14; Length 111;
Best Local Similarity 100.0%; Pred. No. 1.4e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 122 AAAPQPV 128
Db 94 AAAPQPV 100

RESULT 24
US-10-424-599-160298
; Sequence 160298, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 193606
; LENGTH: 106
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_115768C.1.pap
US-10-424-599-160298

Query Match      2.9%; Score 7; DB 12; Length 133;
Best Local Similarity 100.0%; Pred. No. 1.6e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 171 YSYKRGK 177
Db 20 YSYKRGK 26

RESULT 25
US-10-424-599-224277
; Sequence 224277, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 224277
; LENGTH: 134
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_44550C.1.pap
US-10-424-599-224277

Query Match      2.9%; Score 7; DB 12; Length 134;
Best Local Similarity 100.0%; Pred. No. 1.6e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 161 IAIGAGI 167
Db 42 IAIGAGI 48

RESULT 26
US-10-424-599-249048
; Sequence 249048, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 249048
; LENGTH: 137
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_44550C.1.pap
US-10-424-599-249048

```

```

; NAME/KEY: unsure
; LOCATION: (1)..(137)
; OTHER INFORMATION: unsure at all Xaa locations
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_66920C.1.pcp
US-10-424-599-249048

Query Match      2.9%; Score 7; DB 12; Length 137;
Best Local Similarity 100.0%; Pred. No. 1.7e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 171 YSYKRGK 177
   |||||
Db 37 YSYKRGK 43

RESULT 27
US-10-108-260A-3939
; Sequence 3939, Application US/10108260A
; Publication No. US20040005560A1
; GENERAL INFORMATION:
; APPLICANT: HELIX RESEARCH INSTITUTE
; TITLE OF INVENTION: No. US20040005560A1el full length cDNA
; FILE REFERENCE: H1-A0106
; CURRENT APPLICATION NUMBER: US/10/108,260A
; CURRENT FILING DATE: 2002-03-27
; NUMBER OF SEQ ID NOS: 5458
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3939
; LENGTH: 141
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-108-260A-3939

Query Match      2.9%; Score 7; DB 15; Length 141;
Best Local Similarity 100.0%; Pred. No. 1.7e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 40 VSGAGNH 46
   |||||
Db 50 VSGAGNH 56

RESULT 28
US-10-282-122A-59102
; Sequence 59102, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Olsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347

; NAME/KEY: unsure
; LOCATION: (1)..(137)
; OTHER INFORMATION: unsure at all Xaa locations
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_66920C.1.pcp
US-10-424-599-249048

Query Match      2.9%; Score 7; DB 12; Length 137;
Best Local Similarity 100.0%; Pred. No. 1.7e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 171 YSYKRGK 177
   |||||
Db 37 YSYKRGK 43

RESULT 27
US-10-108-260A-3939
; Sequence 3939, Application US/10108260A
; Publication No. US20040005560A1
; GENERAL INFORMATION:
; APPLICANT: HELIX RESEARCH INSTITUTE
; TITLE OF INVENTION: No. US20040005560A1el full length cDNA
; FILE REFERENCE: H1-A0106
; CURRENT APPLICATION NUMBER: US/10/108,260A
; CURRENT FILING DATE: 2002-03-27
; NUMBER OF SEQ ID NOS: 5458
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3939
; LENGTH: 141
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-108-260A-3939

Query Match      2.9%; Score 7; DB 15; Length 141;
Best Local Similarity 100.0%; Pred. No. 1.7e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 40 VSGAGNH 46
   |||||
Db 50 VSGAGNH 56

RESULT 28
US-10-282-122A-59102
; Sequence 59102, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Olsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347

; NAME/KEY: unsure
; LOCATION: (1)..(137)
; OTHER INFORMATION: unsure at all Xaa locations
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_66920C.1.pcp
US-10-424-599-249048

Query Match      2.9%; Score 7; DB 12; Length 137;
Best Local Similarity 100.0%; Pred. No. 1.7e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 171 YSYKRGK 177
   |||||
Db 37 YSYKRGK 43

RESULT 27
US-10-108-260A-3939
; Sequence 3939, Application US/10108260A
; Publication No. US20040005560A1
; GENERAL INFORMATION:
; APPLICANT: HELIX RESEARCH INSTITUTE
; TITLE OF INVENTION: No. US20040005560A1el full length cDNA
; FILE REFERENCE: H1-A0106
; CURRENT APPLICATION NUMBER: US/10/108,260A
; CURRENT FILING DATE: 2002-03-27
; NUMBER OF SEQ ID NOS: 5458
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3939
; LENGTH: 141
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-108-260A-3939

Query Match      2.9%; Score 7; DB 15; Length 141;
Best Local Similarity 100.0%; Pred. No. 1.7e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 40 VSGAGNH 46
   |||||
Db 50 VSGAGNH 56

RESULT 28
US-10-282-122A-59102
; Sequence 59102, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Olsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347

; NAME/KEY: unsure
; LOCATION: (1)..(137)
; OTHER INFORMATION: unsure at all Xaa locations
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_66920C.1.pcp
US-10-424-599-249048

Query Match      2.9%; Score 7; DB 12; Length 137;
Best Local Similarity 100.0%; Pred. No. 1.7e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 171 YSYKRGK 177
   |||||
Db 37 YSYKRGK 43

RESULT 27
US-10-108-260A-3939
; Sequence 3939, Application US/10108260A
; Publication No. US20040005560A1
; GENERAL INFORMATION:
; APPLICANT: HELIX RESEARCH INSTITUTE
; TITLE OF INVENTION: No. US20040005560A1el full length cDNA
; FILE REFERENCE: H1-A0106
; CURRENT APPLICATION NUMBER: US/10/108,260A
; CURRENT FILING DATE: 2002-03-27
; NUMBER OF SEQ ID NOS: 5458
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3939
; LENGTH: 141
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-108-260A-3939

Query Match      2.9%; Score 7; DB 15; Length 141;
Best Local Similarity 100.0%; Pred. No. 1.7e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 40 VSGAGNH 46
   |||||
Db 50 VSGAGNH 56

RESULT 28
US-10-282-122A-59102
; Sequence 59102, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Olsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347

; NAME/KEY: unsure
; LOCATION: (1)..(137)
; OTHER INFORMATION: unsure at all Xaa locations
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_66920C.1.pcp
US-10-424-599-249048

Query Match      2.9%; Score 7; DB 12; Length 137;
Best Local Similarity 100.0%; Pred. No. 1.7e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 171 YSYKRGK 177
   |||||
Db 37 YSYKRGK 43

RESULT 27
US-10-108-260A-3939
; Sequence 3939, Application US/10108260A
; Publication No. US20040005560A1
; GENERAL INFORMATION:
; APPLICANT: HELIX RESEARCH INSTITUTE
; TITLE OF INVENTION: No
```

```
; Sequence 846, Application US/10289762
; Publication No. US20040006218A1
; GENERAL INFORMATION:
; APPLICANT: Griffiths, R.
; TITLE OF INVENTION: Chlamydia pneumoniae genomic sequence and polypeptides, fragments
; TITLE OF INVENTION: thereof and uses thereof, in particular for the diagnosis, prevention
; TITLE OF INVENTION: and treatment of infection
; FILE REFERENCE: 9710-003-999
; CURRENT FILING DATE: 2003-03-27
; CURRENT APPLICATION NUMBER: US/10/289,762
; NUMBER OF SEQ ID NOS: 6849
; SEQ ID NO 846
; LENGTH: 182
; TYPE: PRT
; ORGANISM: Chlamydia pneumoniae
; FEATURE:
; NAME/KEY: SITE
; LOCATION: 1...182
; OTHER INFORMATION: Xaa=unknown or other
US-10-289-762-846

Query Match      2.9%; Score 7; DB 15; Length 182;
Best Local Similarity 100.0%; Pred. No. 2.2e+02; Indels 0; Gaps 0;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 143 KDLGTLG 149
Db 42 KDLGTLG 48

RESULT 31
US-10-424-599-175867
; Sequence 175867, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 175867
; LENGTH: 210
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)...(210)
; OTHER INFORMATION: unsure at all Xaa locations
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_129826C.1.pap
US-10-424-599-175867

Query Match      2.9%; Score 7; DB 12; Length 210;
Best Local Similarity 100.0%; Pred. No. 2.4e+02; Indels 0; Gaps 0;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 161 IAIGAGI 167
Db 181 IAIGAGI 187

RESULT 32
US-10-289-762-849
; Sequence 849, Application US/10289762
; Publication No. US20040006218A1
; GENERAL INFORMATION:
; APPLICANT: Griffiths, R.
; TITLE OF INVENTION: Chlamydia pneumoniae genomic sequence and polypeptides, fragments
```

```
; TITLE OF INVENTION: thereof and uses thereof, in particular for the diagnosis, prevention
; TITLE OF INVENTION: and treatment of infection
; FILE REFERENCE: 9710-003-999
; CURRENT APPLICATION NUMBER: US/10/289,762
; CURRENT FILING DATE: 2003-03-27
; NUMBER OF SEQ ID NOS: 6849
; SEQ ID NO 849
; LENGTH: 228
; TYPE: PRT
; ORGANISM: Chlamydia pneumoniae
US-10-289-762-849

Query Match      2.9%; Score 7; DB 15; Length 228;
Best Local Similarity 100.0%; Pred. No. 2.6e+02; Indels 0; Gaps 0;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 143 KDLGTLG 149
Db 158 KDLGTLG 164

RESULT 33
US-09-764-868-727
; Sequence 727, Application US/09764868
; Patent No. US20020168711A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PT232
; CURRENT APPLICATION NUMBER: US/09/764,868
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - refer to PALM or file wrapper
; NUMBER OF SEQ ID NOS: 1510
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 727
; LENGTH: 237
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (67)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (119)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-764-868-727

Query Match      2.9%; Score 7; DB 9; Length 237;
Best Local Similarity 100.0%; Pred. No. 2.7e+02; Indels 0; Gaps 0;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 128 VIGISQR 134
Db 215 VIGISQR 221

RESULT 34
US-09-764-875-891
; Sequence 891, Application US/09764875
; Publication No. US20040018969A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: P0202
; CURRENT APPLICATION NUMBER: US/09/764,875
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 1249
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 891
; LENGTH: 237
; TYPE: PRT
; ORGANISM: Homo sapiens
```



```
;
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (182)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-764-875-891

Query Match      2.9%; Score 7; DB 11; Length 237;
Best Local Similarity 100.0%; Pred. No. 2.7e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      128 VIGISQR 134
Db      215 VIGISQR 221

RESULT 35
US-09-764-875-1167
; Sequence 1167, Application US/09764875
; Publication No. US20040018969A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PUZ02
; CURRENT APPLICATION NUMBER: US/09/764,875
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 1249
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1167
; LENGTH: 237
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (67)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (119)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-764-875-1167

Query Match      2.9%; Score 7; DB 11; Length 237;
Best Local Similarity 100.0%; Pred. No. 2.7e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      128 VIGISQR 134
Db      215 VIGISQR 221

RESULT 36
US-09-815-242-5680
; Sequence 5680, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
```

```
;
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 5680
; LENGTH: 258
; TYPE: PRT
; ORGANISM: Staphylococcus aureus
US-09-815-242-5680

Query Match      2.9%; Score 7; DB 9; Length 258;
Best Local Similarity 100.0%; Pred. No. 2.9e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      126 QPVIGIS 132
Db      56 QPVIGIS 62

RESULT 37
US-10-424-599-284036
; Sequence 284036, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 284036
; LENGTH: 273
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_9850C.1.pgp
US-10-424-599-284036

Query Match      2.9%; Score 7; DB 12; Length 273;
Best Local Similarity 100.0%; Pred. No. 3.1e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      167 IILGYSY 173
Db      205 IILGYSY 211

RESULT 38
US-10-282-122A-44430
; Sequence 44430, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
```

APPLICANT: Forsyth, R.  
APPLICANT: Xu, H.  
TITLE OF INVENTION: Identification of Essential Genes in Microorganisms  
FILE REFERENCE: ELTRA.034A  
CURRENT APPLICATION NUMBER: US/10/282,122A  
CURRENT FILING DATE: 2003-02-20  
PRIOR APPLICATION NUMBER: 60/191,078  
PRIOR FILING DATE: 2000-03-21  
PRIOR APPLICATION NUMBER: 60/206,848  
PRIOR FILING DATE: 2000-05-23  
PRIOR APPLICATION NUMBER: 60/207,727  
PRIOR FILING DATE: 2000-05-26  
PRIOR APPLICATION NUMBER: 60/230,335  
PRIOR FILING DATE: 2000-09-06  
PRIOR APPLICATION NUMBER: 60/230,347  
PRIOR FILING DATE: 2000-09-09  
PRIOR APPLICATION NUMBER: 60/242,578  
PRIOR FILING DATE: 2000-10-23  
PRIOR APPLICATION NUMBER: 60/253,625  
PRIOR FILING DATE: 2000-11-27  
PRIOR APPLICATION NUMBER: 60/257,931  
PRIOR FILING DATE: 2000-12-22  
PRIOR APPLICATION NUMBER: 60/267,636  
PRIOR FILING DATE: 2001-02-09  
PRIOR APPLICATION NUMBER: 60/269,308  
PRIOR FILING DATE: 2001-02-16  
Remaining Prior Application data removed - See File Wrapper or PALM.  
NUMBER OF SEQ ID NOS: 78614  
SOFTWARE: PatentIn version 3.1  
SEQ ID NO 4430  
LENGTH: 286  
TYPE: PRT  
ORGANISM: Staphylococcus aureus  
US-10-282-122A-4430

Query Match 2.9%; Score 7; DB 12; Length 286;  
Best Local Similarity 100.0%; Pred. No. 3.2e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 126 QPVGIS 132  
DB 59 QPVGIS 65

RESULT 39  
US-09-815-242-12173  
Sequence 12173, Application US/09815242  
Patent No. US20020061569A1  
GENERAL INFORMATION:  
APPLICANT: Haselbeck, Robert  
APPLICANT: Ohlsen, Kari L.  
APPLICANT: Zyskind, Judith W.  
APPLICANT: Wall, Daniel  
APPLICANT: Trawick, John D.  
APPLICANT: Carr, Grant J.  
APPLICANT: Yamamoto, Robert T.  
APPLICANT: Xu, H. Howard  
TITLE OF INVENTION: Identification of Essential Genes in  
FILE REFERENCE: ELTRA.011A  
CURRENT APPLICATION NUMBER: US/09/815,242  
CURRENT FILING DATE: 2001-03-21  
PRIOR APPLICATION NUMBER: 60/191,078  
PRIOR FILING DATE: 2000-03-21  
PRIOR APPLICATION NUMBER: 60/206,848  
PRIOR FILING DATE: 2000-05-23  
PRIOR APPLICATION NUMBER: 60/207,727  
PRIOR FILING DATE: 2000-05-26  
PRIOR APPLICATION NUMBER: 60/242,578  
PRIOR FILING DATE: 2000-10-23  
PRIOR APPLICATION NUMBER: 60/253,625  
PRIOR FILING DATE: 2000-11-27  
PRIOR APPLICATION NUMBER: 60/257,931

PRIOR FILING DATE: 2000-12-22  
PRIOR APPLICATION NUMBER: 60/269,308  
PRIOR FILING DATE: 2001-02-16  
NUMBER OF SEQ ID NOS: 14110  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 12173  
LENGTH: 300  
TYPE: PRT  
ORGANISM: Staphylococcus aureus  
US-09-815-242-12173

Query Match 2.9%; Score 7; DB 9; Length 300;  
Best Local Similarity 100.0%; Pred. No. 3.4e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 126 QPVGIS 132  
DB 73 QPVGIS 79

## RESULT 40

US-10-425-114-61719  
Sequence 61719, Application US/10425114  
Publication No. US20040034888A1  
GENERAL INFORMATION:  
APPLICANT: Liu, Jingdong  
APPLICANT: Zhou, Yihua  
APPLICANT: Kovalic, David K.  
APPLICANT: Screen, Steven E.  
APPLICANT: Tabaska, Jack E.  
APPLICANT: Cao, Yongwei  
TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With  
FILE REFERENCE: 38-21(53313)B  
CURRENT APPLICATION NUMBER: US/10/425,114  
CURRENT FILING DATE: 2003-04-28  
NUMBER OF SEQ ID NOS: 73128  
SEQ ID NO 61719  
LENGTH: 318  
TYPE: PRT  
ORGANISM: Zea mays  
FEATURE:  
OTHER INFORMATION: Clone ID: UC-ZMFLB73162E09\_FLI.pap  
US-10-425-114-61719

Query Match 2.9%; Score 7; DB 12; Length 318;  
Best Local Similarity 100.0%; Pred. No. 3.5e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 118 ARSEAAA 124  
DB 194 ARSEAAA 200

## RESULT 41

US-09-943-002-12  
Sequence 12, Application US/09943002  
Patent No. US20020045734A1  
GENERAL INFORMATION:  
APPLICANT: Durcan, Roy  
TITLE OF INVENTION: NOVEL REOVIRUS-DERIVED PROTEINS AND USES THEREFOR  
FILE REFERENCE: 78973-1C  
CURRENT APPLICATION NUMBER: US/09/943,002  
CURRENT FILING DATE: 2001-08-31  
NUMBER OF SEQ ID NOS: 15  
SOFTWARE: FastSeq for Windows Version 3.0  
SEQ ID NO 12  
LENGTH: 323  
TYPE: PRT  
ORGANISM: Nelson Bay virus  
US-09-943-002-12

Query Match 2.9%; Score 7; DB 9; Length 323;

```
Best Local Similarity 100.0%; Pred. No. 3.6e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 31 AQSGLAS 37
Db 118 AQSGLAS 124

RESULT 42
US-10-369-493-356
; Sequence 356, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 356
; LENGTH: 330
; TYPE: PRT
; ORGANISM: Xenorhabdus nematophilus
US-10-369-493-356

Query Match 2.9%; Score 7; DB 15; Length 330;
Best Local Similarity 100.0%; Pred. No. 3.7e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 161 IAIGAGI 167
Db 193 IAIGAGI 199

RESULT 43
US-10-369-493-21323
; Sequence 21323, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 21323
; LENGTH: 330
; TYPE: PRT
; ORGANISM: Xenorhabdus nematophilus
US-10-369-493-21323

Query Match 2.9%; Score 7; DB 15; Length 330;
Best Local Similarity 100.0%; Pred. No. 3.7e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 161 IAIGAGI 167
Db 193 IAIGAGI 199
```

```
RESULT 44
US-09-815-242-11065
; Sequence 11065, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; TITLE OF INVENTION: Prokaryotes
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 11065
; LENGTH: 335
; TYPE: PRT
; ORGANISM: Haemophilus influenzae
US-09-815-242-11065

Query Match 2.9%; Score 7; DB 9; Length 335;
Best Local Similarity 100.0%; Pred. No. 3.7e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 161 IAIGAGI 167
Db 198 IAIGAGI 204

RESULT 45
US-10-282-122A-58236
; Sequence 58236, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
```

```
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 58236
; LENGTH: 335
; TYPE: PRT
; ORGANISM: Haemophilus influenzae
US-10-282-122A-58236

Query Match          2.9%; Score 7; DB 12; Length 335;
Best Local Similarity 100.0%; Pred. No. 3.7e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 161 IAIGAGI 167
Db 198 IAIGAGI 204

RESULT 46
US-10-408-456-22
; Sequence 22, Application US/10408456
; Publication No. US20040013648A1
; GENERAL INFORMATION:
; APPLICANT: Oxford Biomedica (UK) Limited
; APPLICANT: KINGSMAN, et al., Alan John
; TITLE OF INVENTION: Vector System
; FILE REFERENCE: 674523-2016
; CURRENT APPLICATION NUMBER: US/10/408,456
; CURRENT FILING DATE: 2003-04-08
; PRIOR APPLICATION NUMBER: PCT/GB01/04433
; PRIOR FILING DATE: 2001-10-15
; PRIOR APPLICATION NUMBER: 0024550.6
; PRIOR FILING DATE: 2000-10-06
; NUMBER OF SEQ ID NOS: 34
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 22
; LENGTH: 338
; TYPE: PRT
; ORGANISM: Homo Sapiens
US-10-408-456-22

Query Match          2.9%; Score 7; DB 15; Length 338;
Best Local Similarity 100.0%; Pred. No. 3.7e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 121 EAAAVQP 127
Db 256 EAAAVQP 262

RESULT 47
US-10-408-456-24
; Sequence 24, Application US/10408456
; Publication No. US20040013648A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Karl
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
```

```
; APPLICANT: Oxford Biomedica (UK) Limited
; APPLICANT: KINGSMAN, et al., Alan John
; TITLE OF INVENTION: Vector System
; FILE REFERENCE: 674523-2016
; CURRENT APPLICATION NUMBER: US/10/408,456
; CURRENT FILING DATE: 2003-04-08
; PRIOR APPLICATION NUMBER: PCT/GB01/04433
; PRIOR FILING DATE: 2001-10-15
; PRIOR APPLICATION NUMBER: 0024550.6
; PRIOR FILING DATE: 2000-10-06
; NUMBER OF SEQ ID NOS: 34
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 24
; LENGTH: 338
; TYPE: PRT
; ORGANISM: Homo Sapiens
US-10-408-456-24

Query Match          2.9%; Score 7; DB 15; Length 338;
Best Local Similarity 100.0%; Pred. No. 3.7e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 121 EAAAVQP 127
Db 256 EAAAVQP 262

RESULT 48
US-10-408-456-26
; Sequence 26, Application US/10408456
; Publication No. US20040013648A1
; GENERAL INFORMATION:
; APPLICANT: Oxford Biomedica (UK) Limited
; APPLICANT: KINGSMAN, et al., Alan John
; TITLE OF INVENTION: Vector System
; FILE REFERENCE: 674523-2016
; CURRENT APPLICATION NUMBER: US/10/408,456
; CURRENT FILING DATE: 2003-04-08
; PRIOR APPLICATION NUMBER: PCT/GB01/04433
; PRIOR FILING DATE: 2001-10-15
; PRIOR APPLICATION NUMBER: 0024550.6
; PRIOR FILING DATE: 2000-10-06
; NUMBER OF SEQ ID NOS: 34
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 26
; LENGTH: 338
; TYPE: PRT
; ORGANISM: Homo Sapiens
US-10-408-456-26

Query Match          2.9%; Score 7; DB 15; Length 338;
Best Local Similarity 100.0%; Pred. No. 3.7e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 121 EAAAVQP 127
Db 256 EAAAVQP 262

RESULT 49
US-10-282-122A-57954
; Sequence 57954, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Karl
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
```

```
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.0344
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 57954
; LENGTH: 355
; TYPE: PRT
; ORGANISM: Enterococcus faecium
US-10-282-122A-57954

Query Match          2.9%  Score 7;  DB 12;  Length 355;
Best Local Similarity 100.0%; Pred. No. 3.9e+02;
Matches 7;  Conservative 0;  Mismatches 0;  Indels 0;  Gaps 0;

Qy      160 IIAIGAG 166
      |||||
Db      211 IIAIGAG 217

RESULT 50
US-10-312-273-63
; Sequence 63, Application US/10312273
; Publication No. US20040005667A1
; GENERAL INFORMATION:
; APPLICANT: CHIRON SpA
; TITLE OF INVENTION: IMMUNISATION AGAINST CHLAMYDIA PNEUMONIAE
; FILE REFERENCE: P025035WO
; CURRENT APPLICATION NUMBER: US/10/312,273
; CURRENT FILING DATE: 2002-12-20
; PRIOR APPLICATION NUMBER: 0016363.4
; PRIOR FILING DATE: 2000-07-03
; PRIOR APPLICATION NUMBER: 0017047.2
; PRIOR FILING DATE: 2000-07-11
; PRIOR APPLICATION NUMBER: 0017983.8
; PRIOR FILING DATE: 2000-07-21
; PRIOR APPLICATION NUMBER: 0019368.0
; PRIOR FILING DATE: 2000-08-07
; PRIOR APPLICATION NUMBER: 0020440.4
; PRIOR FILING DATE: 2000-08-18
; PRIOR APPLICATION NUMBER: 0022583.9
; PRIOR FILING DATE: 2000-09-14
; PRIOR APPLICATION NUMBER: 0027549.5
; PRIOR FILING DATE: 2000-11-10
; PRIOR APPLICATION NUMBER: 0031706.5
; PRIOR FILING DATE: 2000-12-22
; NUMBER OF SEQ ID NOS: 664
; SOFTWARE: SeqWin99, version 1.02
; SEQ ID NO 81
; LENGTH: 365
; TYPE: PRT
; ORGANISM: Chlamydia pneumoniae
US-10-312-273-81

Query Match          2.9%  Score 7;  DB 15;  Length 365;
Best Local Similarity 100.0%; Pred. No. 4e+02;
Matches 7;  Conservative 0;  Mismatches 0;  Indels 0;  Gaps 0;

Qy      143 KDLGTLG 149
      |||||
Db      142 KDLGTLG 148

RESULT 52
US-10-425-114-42346
; Sequence 42346, Application US/10425114
; Publication No. US20040034888A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Jingdong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E
; APPLICANT: Tabaska, Jack E
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53313)B
; CURRENT APPLICATION NUMBER: US/10/425,114
```

```
; SEQ ID NO 63
; LENGTH: 356
; TYPE: PRT
; ORGANISM: Chlamydia pneumoniae
US-10-312-273-63

Query Match          2.9%  Score 7;  DB 15;  Length 356;
Best Local Similarity 100.0%; Pred. No. 3.9e+02;
Matches 7;  Conservative 0;  Mismatches 0;  Indels 0;  Gaps 0;

Qy      143 KDLGTLG 149
      |||||
Db      2  KDLGTLG 8

RESULT 51
US-10-312-273-81
; Sequence 81, Application US/10312273
; Publication No. US20040005667A1
; GENERAL INFORMATION:
; APPLICANT: CHIRON SpA
; TITLE OF INVENTION: IMMUNISATION AGAINST CHLAMYDIA PNEUMONIAE
; FILE REFERENCE: P025035WO
; CURRENT APPLICATION NUMBER: US/10/312,273
; CURRENT FILING DATE: 2002-12-20
; PRIOR APPLICATION NUMBER: 0016363.4
; PRIOR FILING DATE: 2000-07-03
; PRIOR APPLICATION NUMBER: 0017047.2
; PRIOR FILING DATE: 2000-07-11
; PRIOR APPLICATION NUMBER: 0017983.8
; PRIOR FILING DATE: 2000-07-21
; PRIOR APPLICATION NUMBER: 0019368.0
; PRIOR FILING DATE: 2000-08-07
; PRIOR APPLICATION NUMBER: 0020440.4
; PRIOR FILING DATE: 2000-08-18
; PRIOR APPLICATION NUMBER: 0022583.9
; PRIOR FILING DATE: 2000-09-14
; PRIOR APPLICATION NUMBER: 0027549.5
; PRIOR FILING DATE: 2000-11-10
; PRIOR APPLICATION NUMBER: 0031706.5
; PRIOR FILING DATE: 2000-12-22
; NUMBER OF SEQ ID NOS: 664
; SOFTWARE: SeqWin99, version 1.02
; SEQ ID NO 81
; LENGTH: 365
; TYPE: PRT
; ORGANISM: Chlamydia pneumoniae
US-10-312-273-81

Query Match          2.9%  Score 7;  DB 15;  Length 365;
Best Local Similarity 100.0%; Pred. No. 4e+02;
Matches 7;  Conservative 0;  Mismatches 0;  Indels 0;  Gaps 0;

Qy      143 KDLGTLG 149
      |||||
Db      142 KDLGTLG 148

RESULT 52
US-10-425-114-42346
; Sequence 42346, Application US/10425114
; Publication No. US20040034888A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Jingdong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E
; APPLICANT: Tabaska, Jack E
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53313)B
; CURRENT APPLICATION NUMBER: US/10/425,114
```

; CURRENT FILING DATE: 2003-04-28  
; NUMBER OF SEQ ID NOS: 73128  
; SEQ ID NO 42346  
; LENGTH: 372  
; TYPE: PRT  
; ORGANISM: Zea mays  
; FEATURE:  
; OTHER INFORMATION: Clone ID: 700209575\_FJI.pap  
US-10-425-114-42346

Query Match 2.9%; Score 7; DB 12; Length 372;  
Best Local Similarity 100.0%; Pred. No. 4.1e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 32 QSGLASA 38  
|||  
DB 290 QSGLASA 296

RESULT 53  
US-09-764-875-779  
; Sequence 779, Application US/09764875  
; Publication No. US20040018969A1  
; GENERAL INFORMATION:  
; APPLICANT: Rosen et al.  
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies  
; FILE REFERENCE: PJ202  
; CURRENT APPLICATION NUMBER: US/09/764,875  
; CURRENT FILING DATE: 2001-01-17  
; Prior application data removed - consult PALM or file wrapper  
; NUMBER OF SEQ ID NOS: 1249  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 779  
; LENGTH: 396  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-764-875-779

Query Match 2.9%; Score 7; DB 11; Length 396;  
Best Local Similarity 100.0%; Pred. No. 4.3e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 122 AAAPQPV 128  
|||  
DB 234 AAAPQPV 240

RESULT 54  
US-09-764-868-850  
; Sequence 850, Application US/09764868  
; Patent No. US20020168711A1  
; GENERAL INFORMATION:  
; APPLICANT: Rosen et al.  
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies  
; FILE REFERENCE: PT232  
; CURRENT APPLICATION NUMBER: US/09/764,868  
; CURRENT FILING DATE: 2001-01-17  
; Prior application data removed - refer to PALM or file wrapper  
; NUMBER OF SEQ ID NOS: 1510  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 850  
; LENGTH: 399  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-764-868-850

Query Match 2.9%; Score 7; DB 9; Length 399;  
Best Local Similarity 100.0%; Pred. No. 4.3e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 122 AAAPQPV 128  
|||  
DB 234 AAAPQPV 240

RESULT 55  
US-10-282-122A-73105  
; Sequence 73105, Application US/10282122A  
; Publication No. US20040029129A1  
; GENERAL INFORMATION:  
; APPLICANT: Wang, Liangsu  
; APPLICANT: Zamudio, Carlos  
; APPLICANT: Malone, Cheryl  
; APPLICANT: Haselbeck, Robert  
; APPLICANT: Chlsen, Kari  
; APPLICANT: Zyskind, Judith  
; APPLICANT: Wall, Daniel  
; APPLICANT: Trawick, John  
; APPLICANT: Carr, Grant  
; APPLICANT: Yamamoto, Robert  
; APPLICANT: Forsyth, R.  
; APPLICANT: Xu, H.  
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms  
; FILE REFERENCE: ELITRA.034A  
; CURRENT APPLICATION NUMBER: US/10/282,122A  
; CURRENT FILING DATE: 2003-02-20  
; PRIOR APPLICATION NUMBER: 60/191,078  
; PRIOR FILING DATE: 2000-03-21  
; PRIOR APPLICATION NUMBER: 60/206,848  
; PRIOR FILING DATE: 2000-05-23  
; PRIOR APPLICATION NUMBER: 60/207,727  
; PRIOR FILING DATE: 2000-05-26  
; PRIOR APPLICATION NUMBER: 60/230,335  
; PRIOR FILING DATE: 2000-09-06  
; PRIOR APPLICATION NUMBER: 60/230,347  
; PRIOR FILING DATE: 2000-09-09  
; PRIOR APPLICATION NUMBER: 60/242,578  
; PRIOR FILING DATE: 2000-10-23  
; PRIOR APPLICATION NUMBER: 60/253,625  
; PRIOR FILING DATE: 2000-11-27  
; PRIOR APPLICATION NUMBER: 60/257,931  
; PRIOR FILING DATE: 2000-12-22  
; PRIOR APPLICATION NUMBER: 60/267,636  
; PRIOR FILING DATE: 2001-02-09  
; PRIOR APPLICATION NUMBER: 60/269,308  
; PRIOR FILING DATE: 2001-02-16  
; Remaining Prior Application data removed - See File Wrapper or PALM.  
; NUMBER OF SEQ ID NOS: 78614  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 73105  
; LENGTH: 419  
; TYPE: PRT  
; ORGANISM: Salmonella paratyphi A  
; FEATURE:  
; NAME/KEY: MISC FEATURE  
; LOCATION: (267)..(267)  
; OTHER INFORMATION: X=any amino acid  
US-10-282-122A-73105

Query Match 2.9%; Score 7; DB 12; Length 419;  
Best Local Similarity 100.0%; Pred. No. 4.5e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 146 GTLGYVL 152  
|||  
DB 182 GTLGYVL 188

RESULT 56  
US-10-282-122A-48132  
; Sequence 48132, Application US/10282122A  
; Publication No. US20040029129A1  
; GENERAL INFORMATION:  
; APPLICANT: Wang, Liangsu  
; APPLICANT: Zamudio, Carlos  
; APPLICANT: Malone, Cheryl

```

; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.0344
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 48132
; LENGTH: 424
; TYPE: PRT
; ORGANISM: Burkholderia cepacia
US-10-282-122A-48132

Query Match          2.9%; Score 7; DB 12; Length 424;
Best Local Similarity 100.0%; Pred. No. 4.6e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      166 GILGYS 172
Db      310 GILGYS 316
      |||||
RESULT 57
US-10-369-493-12020
; Sequence 12020, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 12020
; LENGTH: 432
; TYPE: PRT
; ORGANISM: Mesorhizobium loti

US-10-369-493-12020
Query Match          2.9%; Score 7; DB 15; Length 432;
Best Local Similarity 100.0%; Pred. No. 4.6e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      149 GYVLGIT 155
Db      262 GYVLGIT 268
      |||||
RESULT 58
US-10-369-493-11707
; Sequence 11707, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 11707
; LENGTH: 442
; TYPE: PRT
; ORGANISM: Agrobacterium tumefaciens
US-10-369-493-11707

Query Match          2.9%; Score 7; DB 15; Length 442;
Best Local Similarity 100.0%; Pred. No. 4.7e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      118 ARSEAAA 124
Db      183 ARSEAAA 189
      |||||
RESULT 59
US-10-369-493-12641
; Sequence 12641, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 12641
; LENGTH: 442
; TYPE: PRT
; ORGANISM: Aspergillus nidulans
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(442)
; OTHER INFORMATION: unsure at all Xaa locations
US-10-369-493-12641

Query Match          2.9%; Score 7; DB 15; Length 442;
```

Best Local Similarity 100.0%; Pred. No. 4.7e+02; Mismatches 0; Indels 0; Gaps 0;

QY 33 SGLASAP 39  
 Db 30 SGLASAP 36

## RESULT 60

US-10-282-122A-75534  
 ; Sequence 75534, Application US/10282122A  
 ; Publication No. US20040029129A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Wang, Liangsu  
 ; APPLICANT: Zamudio, Carlos  
 ; APPLICANT: Malone, Cheryl  
 ; APPLICANT: Haselbeck, Robert  
 ; APPLICANT: Ohlsen, Kari  
 ; APPLICANT: Zyskind, Judith  
 ; APPLICANT: Wall, Daniel  
 ; APPLICANT: Trawick, John  
 ; APPLICANT: Carr, Grant  
 ; APPLICANT: Yamamoto, Robert  
 ; APPLICANT: Forsyth, R.  
 ; APPLICANT: Xu, H.  
 ; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms  
 ; FILE REFERENCE: ELITRA.034A  
 ; CURRENT APPLICATION NUMBER: US/10/282,122A  
 ; CURRENT FILING DATE: 2003-02-20  
 ; PRIOR APPLICATION NUMBER: 60/191,078  
 ; PRIOR FILING DATE: 2000-03-21  
 ; PRIOR APPLICATION NUMBER: 60/206,848  
 ; PRIOR FILING DATE: 2000-05-23  
 ; PRIOR APPLICATION NUMBER: 60/207,727  
 ; PRIOR FILING DATE: 2000-05-26  
 ; PRIOR APPLICATION NUMBER: 60/230,335  
 ; PRIOR FILING DATE: 2000-09-06  
 ; PRIOR APPLICATION NUMBER: 60/230,347  
 ; PRIOR FILING DATE: 2000-09-09  
 ; PRIOR APPLICATION NUMBER: 60/242,578  
 ; PRIOR FILING DATE: 2000-10-23  
 ; PRIOR APPLICATION NUMBER: 60/253,625  
 ; PRIOR FILING DATE: 2000-11-27  
 ; PRIOR APPLICATION NUMBER: 60/257,931  
 ; PRIOR FILING DATE: 2000-12-22  
 ; PRIOR APPLICATION NUMBER: 60/267,636  
 ; PRIOR FILING DATE: 2001-02-09  
 ; PRIOR APPLICATION NUMBER: 60/269,308  
 ; PRIOR FILING DATE: 2001-02-16  
 ; REMAINING PRIOR APPLICATION DATA REMOVED - See File Wrapper or PALM.

NUMBER OF SEQ ID NOS: 78614

SOFTWARE: PatentIn version 3.1

SEQ ID NO 75534

LENGTH: 463

TYPE: PRT

ORGANISM: Salmonella typhi

US-10-282-122A-75534

Query Match 2.9%; Score 7; DB 12; Length 463;

Best Local Similarity 100.0%; Pred. No. 4.9e+02; Mismatches 0; Indels 0; Gaps 0;

QY 146 GTLGYVL 152  
 Db 188 GTLGYVL 194

## RESULT 61

US-09-815-242-13961  
 ; Sequence 13961, Application US/09815242  
 ; Patent No. US20020061569A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Haselbeck, Robert

APPLICANT: Ohlsen, Kari L.  
 APPLICANT: Zyskind, Judith W.  
 APPLICANT: Wall, Daniel  
 APPLICANT: Trawick, John D.  
 APPLICANT: Carr, Grant J.  
 APPLICANT: Yamamoto, Robert T.  
 APPLICANT: Xu, H. Howard  
 TITLE OF INVENTION: Identification of Essential Genes in  
 TITLE OF INVENTION: Prokaryotes  
 FILE REFERENCE: ELITRA.011A  
 CURRENT APPLICATION NUMBER: US/09/815,242  
 CURRENT FILING DATE: 2001-03-21  
 PRIOR APPLICATION NUMBER: 60/191,078  
 PRIOR FILING DATE: 2000-03-21  
 PRIOR APPLICATION NUMBER: 60/206,848  
 PRIOR FILING DATE: 2000-05-23  
 PRIOR APPLICATION NUMBER: 60/207,727  
 PRIOR FILING DATE: 2000-05-26  
 PRIOR APPLICATION NUMBER: 60/242,578  
 PRIOR FILING DATE: 2000-10-23  
 PRIOR APPLICATION NUMBER: 60/253,625  
 PRIOR FILING DATE: 2000-11-27  
 PRIOR APPLICATION NUMBER: 60/257,931  
 PRIOR FILING DATE: 2000-12-22  
 PRIOR APPLICATION NUMBER: 60/269,308  
 PRIOR FILING DATE: 2001-02-16  
 NUMBER OF SEQ ID NOS: 14110  
 SOFTWARE: FastSeq for Windows Version 4.0

SEQ ID NO 13961

LENGTH: 465

TYPE: PRT

ORGANISM: Salmonella typhi

FEATURE:

NAME/KEY: VARIANT

LOCATION: (1)...(465)

OTHER INFORMATION: Xaa = Any Amino Acid

US-09-815-242-13961

Query Match 2.9%; Score 7; DB 9; Length 465;

Best Local Similarity 100.0%; Pred. No. 4.9e+02; Mismatches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 146 GTLGYVL 152  
 Db 190 GTLGYVL 196

## RESULT 62

US-10-282-122A-55975  
 ; Sequence 55975, Application US/10282122A  
 ; Publication No. US20040029129A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Wang, Liangsu  
 ; APPLICANT: Zamudio, Carlos  
 ; APPLICANT: Malone, Cheryl  
 ; APPLICANT: Haselbeck, Robert  
 ; APPLICANT: Ohlsen, Kari  
 ; APPLICANT: Zyskind, Judith  
 ; APPLICANT: Wall, Daniel  
 ; APPLICANT: Trawick, John  
 ; APPLICANT: Carr, Grant  
 ; APPLICANT: Yamamoto, Robert  
 ; APPLICANT: Forsyth, R.  
 ; APPLICANT: Xu, H.

TITLE OF INVENTION: Identification of Essential Genes in Microorganisms

FILE REFERENCE: ELITRA.034A

CURRENT APPLICATION NUMBER: US/10/282,122A

CURRENT FILING DATE: 2003-02-20

PRIOR APPLICATION NUMBER: 60/191,078

PRIOR FILING DATE: 2000-03-21

PRIOR APPLICATION NUMBER: 60/206,848

PRIOR FILING DATE: 2000-05-23

PRIOR APPLICATION NUMBER: 60/207,727



```
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 55975
; LENGTH: 465
; TYPE: PRT
; ORGANISM: Enterobacter cloacae
US-10-282-122A-55975

Query Match      2.9%; Score 7; DB 12; Length 465;
Best Local Similarity 100.0%; Pred.No. 4.9e+02; Indels 0; Gaps 0;
Matches 7; Conservative 0; Mismatches 0;

QY 146 GTLGYVL 152
Db 190 GTLGYVL 196

RESULT 63
US-10-282-122A-43395
; Sequence 43395, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 55975
; LENGTH: 465
; TYPE: PRT
; ORGANISM: Enterobacter cloacae
US-10-282-122A-55975

Query Match      2.9%; Score 7; DB 12; Length 465;
Best Local Similarity 100.0%; Pred.No. 4.9e+02; Indels 0; Gaps 0;
Matches 7; Conservative 0; Mismatches 0;

QY 146 GTLGYVL 152
Db 190 GTLGYVL 196

RESULT 64
US-10-282-122A-50382
; Sequence 50382, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 50382
; LENGTH: 469
; TYPE: PRT
; ORGANISM: Burkholderia mallei
US-10-282-122A-50382

Query Match      2.9%; Score 7; DB 12; Length 469;
Best Local Similarity 100.0%; Pred.No. 5e+02; Indels 0; Gaps 0;
Matches 7; Conservative 0; Mismatches 0;
```

```
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 43395
; LENGTH: 469
; TYPE: PRT
; ORGANISM: Klebsiella pneumoniae
US-10-282-122A-43395

Query Match      2.9%; Score 7; DB 12; Length 469;
Best Local Similarity 100.0%; Pred.No. 5e+02; Indels 0; Gaps 0;
Matches 7; Conservative 0; Mismatches 0;

QY 18 SPAPGLR 24
Db 187 SPAPGLR 193

RESULT 64
US-10-282-122A-50382
; Sequence 50382, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 50382
; LENGTH: 469
; TYPE: PRT
; ORGANISM: Burkholderia mallei
US-10-282-122A-50382

Query Match      2.9%; Score 7; DB 12; Length 469;
Best Local Similarity 100.0%; Pred.No. 5e+02; Indels 0; Gaps 0;
Matches 7; Conservative 0; Mismatches 0;
```

```
QY 166 GILIGYS 172
DB 309 GILIGYS 315

RESULT 65
US-10-282-122A-60099
; Sequence 60099, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 60099
; LENGTH: 469
; TYPE: PRT
; ORGANISM: Klebsiella pneumoniae
US-10-282-122A-60099

Query Match 2.9%; Score 7; DB 12; Length 469;
Best Local Similarity 100.0%; Pred. No. 5e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 18 SPAPGLR 24
DB 187 SPAPGLR 193

RESULT 66
US-10-408-456-28
; Sequence 28, Application US/10408456
; Publication No. US20040013648A1
; GENERAL INFORMATION:
; APPLICANT: Oxford Biomedica (UK) Limited
; APPLICANT: KINGSMAN, et al., Alan John
; TITLE OF INVENTION: Vector System
; FILE REFERENCE: 674523-2016

US-10-057-951-2_copy_22_263.oligo.rapb

; CURRENT APPLICATION NUMBER: US/10/408,456
; CURRENT FILING DATE: 2003-04-08
; PRIOR APPLICATION NUMBER: PCT/GB01/04433
; PRIOR FILING DATE: 2001-10-15
; PRIOR APPLICATION NUMBER: 0024550.6
; PRIOR FILING DATE: 2000-10-06
; NUMBER OF SEQ ID NOS: 34
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 28
; LENGTH: 501
; TYPE: PRT
; ORGANISM: Homo Sapiens
US-10-408-456-28

Query Match 2.9%; Score 7; DB 15; Length 501;
Best Local Similarity 100.0%; Pred. No. 5.3e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 121 EAAAVQP 127
DB 419 EAAAVQP 425

RESULT 67
US-10-156-761-11459
; Sequence 11459, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 11459
; LENGTH: 545
; TYPE: PRT
; ORGANISM: Streptomyces avermitilis
US-10-156-761-11459

Query Match 2.9%; Score 7; DB 14; Length 545;
Best Local Similarity 100.0%; Pred. No. 5.7e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 111 APANALP 117
DB 90 APANALP 96

RESULT 68
US-10-425-114-67035
; Sequence 67035, Application US/10425114
; Publication No. US20040034888A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Jingdong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E
; APPLICANT: Tabaska, Jack E
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53313)B
```

; CURRENT APPLICATION NUMBER: US/10/425,114;  
; CURRENT FILING DATE: 2003-04-28  
; NUMBER OF SEQ ID NOS: 73128  
; SEQ ID NO 67035  
; LENGTH: 552  
; TYPE: PRT  
; ORGANISM: Zea mays  
; FEATURE:  
; OTHER INFORMATION: Clone ID: LIB4757-014-F9\_FLI.pbp  
US-10-425-114-67035

Query Match 2.9%; Score 7; DB 12; Length 552;  
Best Local Similarity 100.0%; Pred. No. 5.8e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 18 SPAPGLR 24  
| | | | |  
Db 92 SPAPGLR 98

## RESULT 69

US-09-738-626-5317  
; Sequence 5317, Application US/09738626  
; Publication No. US20020197605A1  
; GENERAL INFORMATION:  
; APPLICANT: NAKAGAWA, SATOSHI  
; APPLICANT: MIZOGUCHI, HIROSHI  
; APPLICANT: ANDO, SEIKO  
; APPLICANT: HAYASHI, MIKIRO  
; APPLICANT: OCHIAI, KEIKO  
; APPLICANT: YOKOI, HARUHIKO  
; APPLICANT: TATEISHI, NAKO  
; APPLICANT: SENOH, AKIHIRO  
; APPLICANT: IKEDA, MASATO  
; APPLICANT: OZAKI, AKIO  
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES  
; FILE REFERENCE: 249-125  
; CURRENT APPLICATION NUMBER: US/09/738,626  
; CURRENT FILING DATE: 2000-12-18  
; PRIOR APPLICATION NUMBER: JP 99/377484  
; PRIOR FILING DATE: 1999-12-16  
; PRIOR APPLICATION NUMBER: JP 00/159162  
; PRIOR FILING DATE: 2000-04-07  
; PRIOR APPLICATION NUMBER: JP 00/280988  
; PRIOR FILING DATE: 2000-08-03  
; NUMBER OF SEQ ID NOS: 7059  
; SOFTWARE: PatentIn ver. 3.0  
; SEQ ID NO 5317  
; LENGTH: 562  
; TYPE: PRT  
; ORGANISM: Corynebacterium glutamicum  
US-09-738-626-5317

Query Match 2.9%; Score 7; DB 9; Length 562;  
Best Local Similarity 100.0%; Pred. No. 5.9e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 145 LGTLGYV 151  
| | | | |  
Db 511 LGTLGYV 517

## RESULT 70

US-10-424-599-146903  
; Sequence 146903, Application US/10424599  
; Publication No. US20040031072A1  
; GENERAL INFORMATION:  
; APPLICANT: La Rosa Thomas J  
; APPLICANT: Kovalic David K  
; APPLICANT: Zhou Yihua  
; APPLICANT: Cao Yongwei  
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With  
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement

; FILE REFERENCE: 38-21(53223)B  
; CURRENT APPLICATION NUMBER: US/10/424,599  
; CURRENT FILING DATE: 2003-04-28  
; NUMBER OF SEQ ID NOS: 285684  
; SEQ ID NO 146903  
; LENGTH: 563  
; TYPE: PRT  
; ORGANISM: Glycine max  
; FEATURE:  
; OTHER INFORMATION: Clone ID: PAT\_MRT3847\_103674C.1.pbp  
US-10-424-599-146903

Query Match 2.9%; Score 7; DB 12; Length 563;  
Best Local Similarity 100.0%; Pred. No. 5.9e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 236 QAGTPGA 242  
| | | | |  
Db 279 QAGTPGA 285

## RESULT 71

US-10-369-493-19493  
; Sequence 19493, Application US/10369493  
; Publication No. US20030233675A1  
; GENERAL INFORMATION:  
; APPLICANT: Cao, Yongwei  
; APPLICANT: Hinkle, Gregory J.  
; APPLICANT: Slater, Steven C.  
; APPLICANT: Goldman, Barry S.  
; APPLICANT: Cher, Xianfeng  
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF  
; TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES  
; FILE REFERENCE: 38-10(52052)B  
; CURRENT APPLICATION NUMBER: US/10/369,493  
; CURRENT FILING DATE: 2003-02-28  
; PRIOR APPLICATION NUMBER: US 60/360,039  
; PRIOR FILING DATE: 2002-02-21  
; NUMBER OF SEQ ID NOS: 47374  
; SEQ ID NO 19493  
; LENGTH: 575  
; TYPE: PRT  
; ORGANISM: Myxococcus xanthus  
US-10-369-493-19493

Query Match 2.9%; Score 7; DB 15; Length 575;  
Best Local Similarity 100.0%; Pred. No. 6e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 117 PARSEAA 123  
| | | | |  
Db 2 PARSEAA 8

## RESULT 72

US-09-946-678-2  
; Sequence 2, Application US/09946678  
; Patent No. US20020106782A1  
; GENERAL INFORMATION:  
; APPLICANT: ITO, Kotaro  
; APPLICANT: UMITSUKI, Genryou  
; APPLICANT: KOYAMA, Yasuji  
; TITLE OF INVENTION: Protein Having Glutaminase Activity and Gene Encoding the Same  
; FILE REFERENCE: 0283-0158P  
; CURRENT APPLICATION NUMBER: US/09/946,678  
; CURRENT FILING DATE: 2001-09-06  
; PRIOR APPLICATION NUMBER: JP 2000-270371  
; PRIOR FILING DATE: 2000-09-06  
; NUMBER OF SEQ ID NOS: 27  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 2  
; LENGTH: 584  
; TYPE: PRT

; ORGANISM: Cryptococcus nodaensis  
US-09-946-678-2

Query Match 2.9%; Score 7; DB 9; Length 684;  
Best Local Similarity 100.0%; Pred. No. 7e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 30 DAQSGLA 36  
Db 411 DAQSGLA 417

## RESULT 73

US-09-815-242-13411  
; Sequence 13411, Application US/09815242  
; Patent No. US20020061569A1  
; GENERAL INFORMATION:  
; APPLICANT: Haselbeck, Robert  
; APPLICANT: Ohlsen, Karl L.  
; APPLICANT: Zyskind, Judith W.  
; APPLICANT: Wall, Daniel  
; APPLICANT: Trawick, John D.  
; APPLICANT: Carr, Grant J.  
; APPLICANT: Yamamoto, Robert T.  
; APPLICANT: Xu, H. Howard

; TITLE OF INVENTION: Identification of Essential Genes in

; FILE REFERENCE: ELITRA 011A

; CURRENT APPLICATION NUMBER: US/09/815,242

; CURRENT FILING DATE: 2001-03-21

; PRIOR APPLICATION NUMBER: 60/191,078

; PRIOR FILING DATE: 2000-03-21

; PRIOR APPLICATION NUMBER: 60/206,848

; PRIOR FILING DATE: 2000-05-23

; PRIOR APPLICATION NUMBER: 60/207,727

; PRIOR FILING DATE: 2000-05-26

; PRIOR APPLICATION NUMBER: 60/242,578

; PRIOR FILING DATE: 2000-10-23

; PRIOR APPLICATION NUMBER: 60/253,625

; PRIOR FILING DATE: 2000-11-27

; PRIOR APPLICATION NUMBER: 60/257,931

; PRIOR FILING DATE: 2000-12-22

; PRIOR APPLICATION NUMBER: 60/269,308

; PRIOR FILING DATE: 2001-02-16

; NUMBER OF SEQ ID NOS: 14110

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 13411

; LENGTH: 701

; TYPE: PRT

; ORGANISM: Streptococcus pneumoniae

US-09-815-242-13411

Query Match 2.9%; Score 7; DB 9; Length 701;  
Best Local Similarity 100.0%; Pred. No. 7.1e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 94 EIQEASE 100  
Db 101 EIQEASE 107

## RESULT 74

US-10-282-122A-73782  
; Sequence 73782, Application US/10282122A  
; Publication No. US20040029129A1

; GENERAL INFORMATION:

; APPLICANT: Wang, Liangsu

; APPLICANT: Zamudio, Carlos

; APPLICANT: Malone, Cheryl

; APPLICANT: Haselbeck, Robert

; APPLICANT: Ohlsen, Karl

; APPLICANT: Zyskind, Judith

; APPLICANT: Wall, Daniel

; APPLICANT: Trawick, John  
; APPLICANT: Carr, Grant  
; APPLICANT: Yamamoto, Robert  
; APPLICANT: Forsyth, R.  
; APPLICANT: Xu, H.  
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms  
; FILE REFERENCE: ELITRA 034A  
; CURRENT APPLICATION NUMBER: US/10/282,122A  
; CURRENT FILING DATE: 2003-02-20  
; PRIOR APPLICATION NUMBER: 60/191,078  
; PRIOR FILING DATE: 2000-03-21  
; PRIOR APPLICATION NUMBER: 60/206,848  
; PRIOR FILING DATE: 2000-05-23  
; PRIOR APPLICATION NUMBER: 60/207,727  
; PRIOR FILING DATE: 2000-05-26  
; PRIOR APPLICATION NUMBER: 60/230,335  
; PRIOR FILING DATE: 2000-09-06  
; PRIOR APPLICATION NUMBER: 60/230,347  
; PRIOR FILING DATE: 2000-09-09  
; PRIOR APPLICATION NUMBER: 60/242,578  
; PRIOR FILING DATE: 2000-10-23  
; PRIOR APPLICATION NUMBER: 60/253,625  
; PRIOR FILING DATE: 2000-11-27  
; PRIOR APPLICATION NUMBER: 60/257,931  
; PRIOR FILING DATE: 2000-12-22  
; PRIOR APPLICATION NUMBER: 60/267,636  
; PRIOR FILING DATE: 2001-02-09  
; PRIOR APPLICATION NUMBER: 60/269,308  
; PRIOR FILING DATE: 2001-02-16  
; Remaining Prior Application data removed - See File Wrapper or PALM.  
; NUMBER OF SEQ ID NOS: 78614  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 73782  
; LENGTH: 701  
; TYPE: PRT  
; ORGANISM: Streptococcus pneumoniae  
US-10-282-122A-73782

Query Match 2.9%; Score 7; DB 12; Length 701;  
Best Local Similarity 100.0%; Pred. No. 7.1e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 94 EIQEASE 100  
Db 101 EIQEASE 107

## RESULT 75

US-09-738-626-6019  
; Sequence 6019, Application US/09738626  
; Publication No. US20020197605A1

; GENERAL INFORMATION:

; APPLICANT: NAKAGAWA, SATOSHI

; APPLICANT: MIZOGUCHI, HIROSHI

; APPLICANT: ANDO, SEIKO

; APPLICANT: HAYASHI, MIKIRO

; APPLICANT: OCHIAI, KEIKO

; APPLICANT: YOKOI, HARUHIKO

; APPLICANT: TATEISHI, NAKO

; APPLICANT: SENO, AKIHIRO

; APPLICANT: IKEDA, MASATO

; APPLICANT: OZAKI, AKIO

; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES

; FILE REFERENCE: 249-125

; CURRENT APPLICATION NUMBER: US/09/738,626

; CURRENT FILING DATE: 2000-12-18

; PRIOR APPLICATION NUMBER: JP 99/377484

; PRIOR FILING DATE: 1999-12-16

; PRIOR APPLICATION NUMBER: JP 00/159162

; PRIOR FILING DATE: 2000-04-07

; PRIOR APPLICATION NUMBER: JP 00/280988

; PRIOR FILING DATE: 2000-08-03

; NUMBER OF SEQ ID NOS: 7059

```
; SOFTWARE: PatentIn ver. 3.0
; SEQ ID NO 6019
; LENGTH: 706
; TYPE: PRT
; ORGANISM: Corynebacterium glutamicum
US-09-738-626-6019

Query Match      2.9%  Score 7;  DB 9;  Length 706;
Best Local Similarity 100.0%;  Pred. No. 7.2e+02;
Matches 7;  Conservative 0;  Mismatches 0;  Indels 0;  Gaps 0;

Qy      113 ANALPAR 119
      |||||
Db      608 ANALPAR 614

Search completed: March 18, 2004, 13:25:40
Job time : 44 secs
```



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 17, 2004, 07:23:40 ; Search time 22 Seconds  
(without alignments)  
567.886 Million cell updates/sec

Title: US-10-057-951-2\_COPY\_22\_263  
Perfect score: 242  
Sequence: 1 SGCCFWNGHLYREDTSPA.....PVDPOEGSTPLMGQATPGA 242

Scoring table: OLIGO  
Gapop 60.0 , Gapext 60.0  
Searched: 389414 seqs, 51625971 residues  
Word size : 0  
Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Listing first 100 summaries

Database : Issued Patents AA: \*  
1: /cgn2\_6/ptodata/2/iaa/5A-COMB.pep.\*  
2: /cgn2\_6/ptodata/2/iaa/5B-COMB.pep.\*  
3: /cgn2\_6/ptodata/2/iaa/6A-COMB.pep.\*  
4: /cgn2\_6/ptodata/2/iaa/6B-COMB.pep.\*  
5: /cgn2\_6/ptodata/2/iaa/6C-COMB.pep.\*  
6: /cgn2\_6/ptodata/2/iaa/6D-COMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	242	100.0	263	4	US-09-411-977-2
2	8	3.3	716	2	US-08-766-982-1
3	8	3.3	716	3	US-09-296-219-1
4	7	2.9	182	4	US-03-198-452A-846
5	7	2.9	204	4	US-03-107-532A-5978
6	7	2.9	228	4	US-03-198-452A-849
7	7	2.9	274	4	US-09-134-000C-3673
8	7	2.9	341	4	US-09-252-991A-18308
9	7	2.9	348	4	US-09-489-039A-8827
10	7	2.9	353	4	US-09-107-532A-5053
11	7	2.9	356	4	US-09-107-532A-4004
12	7	2.9	356	4	US-09-107-532A-6936
13	7	2.9	381	4	US-08-252-991A-29970
14	7	2.9	413	4	US-09-252-991A-28999
15	7	2.9	472	4	US-09-489-039A-13479
16	7	2.9	497	1	US-08-009-075-4
17	7	2.9	517	4	US-09-489-039A-10153
18	7	2.9	569	2	US-08-467-822-27
19	7	2.9	569	3	US-08-432-697-27
20	7	2.9	569	3	US-08-466-248-27
21	7	2.9	569	4	US-03-543-681A-6029
22	7	2.9	625	2	US-08-532-547-9
23	7	2.9	625	2	US-08-532-547-9
24	7	2.9	625	3	US-09-019-809-7
25	7	2.9	625	3	US-09-019-809-9
26	7	2.9	625	4	US-09-471-177-7
27	7	2.9	625	4	US-09-471-177-9

28	7	2.9	684	4	US-09-946-678-2	Sequence 2, Appli
29	7	2.9	800	2	US-08-469-537A-72	Sequence 72, Appl
30	7	2.9	800	2	US-08-469-537A-78	Sequence 78, Appl
31	7	2.9	831	4	US-09-269-861A-8	Sequence 8, Appli
32	7	2.9	937	2	US-08-469-537A-105	Sequence 105, App
33	7	2.9	1156	4	US-09-002-285-72	Sequence 72, Appl
34	7	2.9	1156	4	US-09-589-477-72	Sequence 72, Appl
35	7	2.9	1156	4	US-09-661-322A-28	Sequence 28, Appl
36	7	2.9	1157	2	US-08-532-547-5	Sequence 5, Appli
37	7	2.9	1157	2	US-08-379-656B-5	Sequence 5, Appli
38	7	2.9	1157	3	US-08-455-838-5	Sequence 5, Appli
39	7	2.9	1157	3	US-09-019-809-5	Sequence 5, Appli
40	7	2.9	1157	4	US-09-471-177-5	Sequence 5, Appli
41	7	2.9	1247	4	US-09-252-991A-32960	Sequence 32960, A
42	6	2.5	7	4	US-09-192-012-8	Sequence 8, Appli
43	6	2.5	8	1	US-08-267-092A-12	Sequence 12, Appl
44	6	2.5	8	3	US-08-468-161-12	Sequence 12, Appl
45	6	2.5	8	5	PCT-US95-08156-12	Sequence 12, Appl
46	6	2.5	38	4	US-09-348-953-4	Sequence 4, Appli
47	6	2.5	39	4	US-09-227-357-589	Sequence 589, App
48	6	2.5	46	4	US-09-348-953-6	Sequence 6, Appli
49	6	2.5	53	4	US-09-227-357-588	Sequence 588, App
50	6	2.5	59	4	US-08-630-915A-215	Sequence 215, App
51	6	2.5	63	4	US-09-489-039A-10500	Sequence 10500, A
52	6	2.5	66	4	US-09-107-532A-6727	Sequence 6727, Ap
53	6	2.5	72	4	US-03-134-000C-4771	Sequence 4771, Ap
54	6	2.5	73	4	US-09-489-039A-9351	Sequence 9351, Ap
55	6	2.5	73	4	US-09-107-532A-4213	Sequence 4213, Ap
56	6	2.5	78	2	US-08-612-788-13	Sequence 13, Appl
57	6	2.5	78	2	US-08-612-788-14	Sequence 14, Appl
58	6	2.5	78	2	US-08-612-788-15	Sequence 15, Appl
59	6	2.5	78	2	US-08-612-788-16	Sequence 16, Appl
60	6	2.5	78	2	US-08-612-788-17	Sequence 17, Appl
61	6	2.5	78	2	US-08-612-788-18	Sequence 18, Appl
62	6	2.5	78	2	US-08-612-788-19	Sequence 19, Appl
63	6	2.5	78	2	US-08-612-788-20	Sequence 20, Appl
64	6	2.5	78	2	US-08-612-788-21	Sequence 21, Appl
65	6	2.5	78	2	US-08-612-788-22	Sequence 22, Appl
66	6	2.5	78	2	US-08-612-788-23	Sequence 23, Appl
67	6	2.5	78	2	US-08-763-528A-3	Sequence 3, Appli
68	6	2.5	78	2	US-08-763-528A-4	Sequence 4, Appli
69	6	2.5	78	2	US-08-763-528A-5	Sequence 5, Appli
70	6	2.5	78	3	US-09-066-028-13	Sequence 13, Appl
71	6	2.5	78	3	US-09-066-028-14	Sequence 14, Appl
72	6	2.5	78	3	US-09-066-028-15	Sequence 15, Appl
73	6	2.5	78	3	US-09-066-028-16	Sequence 16, Appl
74	6	2.5	78	3	US-09-066-028-17	Sequence 17, Appl
75	6	2.5	78	3	US-09-066-028-18	Sequence 18, Appl
76	6	2.5	78	3	US-09-066-028-19	Sequence 19, Appl
77	6	2.5	78	3	US-09-066-028-20	Sequence 20, Appl
78	6	2.5	78	3	US-09-066-028-21	Sequence 21, Appl
79	6	2.5	78	3	US-09-066-028-22	Sequence 22, Appl
80	6	2.5	78	3	US-09-066-028-23	Sequence 23, Appl
81	6	2.5	78	4	US-09-335-325-13	Sequence 13, Appl
82	6	2.5	78	4	US-09-335-325-14	Sequence 14, Appl
83	6	2.5	78	4	US-09-335-325-15	Sequence 15, Appl
84	6	2.5	78	4	US-09-335-325-16	Sequence 16, Appl
85	6	2.5	78	4	US-09-335-325-17	Sequence 17, Appl
86	6	2.5	78	4	US-09-335-325-18	Sequence 18, Appl
87	6	2.5	78	4	US-09-335-325-19	Sequence 19, Appl
88	6	2.5	78	4	US-09-335-325-20	Sequence 20, Appl
89	6	2.5	78	4	US-09-335-325-21	Sequence 21, Appl
90	6	2.5	78	4	US-09-335-325-22	Sequence 22, Appl
91	6	2.5	78	4	US-09-335-325-23	Sequence 23, Appl
92	6	2.5	78	4	US-09-348-953-3	Sequence 3, Appli
93	6	2.5	79	2	US-08-612-788-7	Sequence 7, Appli
94	6	2.5	79	2	US-08-612-788-8	Sequence 8, Appli
95	6	2.5	79	2	US-08-612-788-9	Sequence 9, Appli
96	6	2.5	79	2	US-08-612-788-10	Sequence 10, Appl
97	6	2.5	79	2	US-08-612-788-11	Sequence 11, Appl
98	6	2.5	79	2	US-08-763-528A-1	Sequence 1, Appli
99	6	2.5	79	2	US-08-763-528A-2	Sequence 2, Appli
100	6	2.5	79	3	US-09-066-028-7	Sequence 7, Appli

# ALIGNMENTS

```

RESULT 1
US-09-411-977-2
; Sequence 2, Application US/09411977
; Patent No. 6372473
; GENERAL INFORMATION:
; APPLICANT: Moore, Paul A.
; APPLICANT: Ruben, Steven M.
; APPLICANT: Ebner, Reinhard
; TITLE OF INVENTION: Tissue Plasminogen Activator-Like Protease
; FILE REFERENCE: P378P1
; CURRENT APPLICATION NUMBER: US/09/411,977
; CURRENT FILING DATE: 1999-10-04
; EARLIER APPLICATION NUMBER: 09/084,491
; EARLIER FILING DATE: 1998-05-27
; EARLIER APPLICATION NUMBER: 60/048,000
; EARLIER FILING DATE: 1997-05-28
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 263
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-411-977-2

Query Match      100.0%; Score 242; DB 4; Length 263;
Best Local Similarity 100.0%; Pred. No. 2.1e-228;
Matches 242; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SGGCFWDNGHLYREDQTSAPGLRCLNWLDAQSLASAPVSGAGNHSYCRNPDEDPGPW 60
DB 22 SGGCFWDNGHLYREDQTSAPGLRCLNWLDAQSLASAPVSGAGNHSYCRNPDEDPGPW 81

QY 61 CVVSGEAGVPEKPCEDLRCPTTQALPAFTTETOASEGPGADEVOVFAPALPARS 120
DB 82 CVVSGEAGVPEKPCEDLRCPTTQALPAFTTETOASEGPGADEVOVFAPALPARS 141

QY 121 EAAAVQPVIGISQVRMNSKEKKDLGTGYVLGITMVIITAIAGAGIILGYXYRGKDLK 180
DB 142 EAAAVQPVIGISQVRMNSKEKKDLGTGYVLGITMVIITAIAGAGIILGYXYRGKDLK 201

QY 181 EHQDKVCEREMQRTLPLSATNTCEIVDEKTVVHTSQTVPDPOEGSTPLMGQAGTP 240
DB 202 EHQDKVCEREMQRTLPLSATNTCEIVDEKTVVHTSQTVPDPOEGSTPLMGQAGTP 261

QY 241 GA 242
DB 262 GA 263

```

```

RESULT 2
US-08-766-982-1
; Sequence 1, Application US/08766982
; Patent No. 5948892
; GENERAL INFORMATION:
; APPLICANT: Wahl, Robert C.
; TITLE OF INVENTION: Analogs of Macrophage Stimulating
; TITLE OF INVENTION: Protein
; NUMBER OF SEQUENCES: 13
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Amgen Inc.
; STREET: 1840 Dehavilland Drive
; CITY: Thousand Oaks
; STATE: California
; COUNTRY: USA
; ZIP: 91320-1789
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/296,219
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Winter, Robert B.
; REFERENCE/DOCKET NUMBER: A-441
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 716 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-766-982-1

Query Match      3.3%; Score 8; DB 3; Length 716;
Best Local Similarity 100.0%; Pred. No. 1.9;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 55 DPRGPWCY 62
DB 163 DPRGPWCY 170

```

```

; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/766,982
; FILING DATE:
; CLASSIFICATION: 536
; ATTORNEY/AGENT INFORMATION:
; NAME: Winter, Robert B.
; REFERENCE/DOCKET NUMBER: A-441
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 716 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-766-982-1

```

```

Query Match      3.3%; Score 8; DB 2; Length 716;
Best Local Similarity 100.0%; Pred. No. 1.9;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 55 DPRGPWCY 62
DB 163 DPRGPWCY 170

```

```

RESULT 3
US-09-296-219-1
; Sequence 1, Application US/09296219
; Patent No. 6248560
; GENERAL INFORMATION:
; APPLICANT: Wahl, Robert C.
; TITLE OF INVENTION: Analogs of Macrophage Stimulating
; TITLE OF INVENTION: Protein
; NUMBER OF SEQUENCES: 13
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Amgen Inc.
; STREET: 1840 Dehavilland Drive
; CITY: Thousand Oaks
; STATE: California
; COUNTRY: USA
; ZIP: 91320-1789
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/296,219
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Winter, Robert B.
; REFERENCE/DOCKET NUMBER: A-441
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 716 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-09-296-219-1

```

```

Query Match      3.3%; Score 8; DB 3; Length 716;
Best Local Similarity 100.0%; Pred. No. 1.9;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 55 DPRGPWCY 62
DB 163 DPRGPWCY 170

```



```
RESULT 4
US-09-198-452A-846
; Sequence 846, Application US/09198452A
; Patent No. 6559294
; GENERAL INFORMATION:
; APPLICANT: Griffiths, R.
; TITLE OF INVENTION: Chlamydia pneumoniae genomic sequence and polypeptides, fragments
; TITLE OF INVENTION: thereof and uses thereof, in particular for the diagnosis, prevention
; TITLE OF INVENTION: and treatment of infection
; FILE REFERENCE: 9710-003-999
; CURRENT APPLICATION NUMBER: US/09/198,452A
; CURRENT FILING DATE: 1998-11-24
; NUMBER OF SEQ ID NOS: 6849
; SEQ ID NO 846
; LENGTH: 182
; TYPE: PRT
; ORGANISM: Chlamydia pneumoniae
; FEATURE:
; NAME/KEY: SITE
; LOCATION: 1...182
; OTHER INFORMATION: Xaa=unknown or other
US-09-198-452A-846

Query Match          2.9%; Score 7; DB 4; Length 182;
Best Local Similarity 100.0%; Pred. No. 51;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      143 KDLGTLG 149
      |||||
Db      42 KDLGTLG 48

RESULT 5
US-09-107-532A-5978
; Sequence 5978, Application US/09107532A
; Patent No. 6583275
; GENERAL INFORMATION:
; APPLICANT: Lynn A Doucette-Stamm and David Bush
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: GENOME THERAPEUTICS CORPORATION
; STREET: 100 Beaver Street
; CITY: Waltham
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02354
; COMPUTER READABLE FORM:
; MEDIUM TYPE: CD-ROM ISO9660
; COMPUTER: PC
; OPERATING SYSTEM: <Unknown>
; SOFTWARE: ASCII
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/107,532A
; FILING DATE: 30-Jun-1998
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/085,598
; FILING DATE: 14 May 1998
; APPLICATION NUMBER: 60/051571
; FILING DATE: July 2, 1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Ariniello, Pamela Denise
; REGISTRATION NUMBER: 40,489
; REFERENCE/DOCKET NUMBER: GTC-012
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (781)893-5007
; TELEFAX: (781)893-8277
; INFORMATION FOR SEQ ID NO: 5978:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 204 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
```

```
; MOLECULE TYPE: protein
; HYPOTHETICAL: YES
; ORIGINAL SOURCE:
; ORGANISM: Enterococcus faecium
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (B) LOCATION 1...204
; SEQUENCE DESCRIPTION: SEQ ID NO: 5978:
US-09-107-532A-5978

Query Match          2.9%; Score 7; DB 4; Length 204;
Best Local Similarity 100.0%; Pred. No. 57;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      33 SGLASAP 39
      |||||
Db      98 SGLASAP 104

RESULT 6
US-09-198-452A-849
; Sequence 849, Application US/09198452A
; Patent No. 6559294
; GENERAL INFORMATION:
; APPLICANT: Griffiths, R.
; TITLE OF INVENTION: Chlamydia pneumoniae genomic sequence and polypeptides, fragments
; TITLE OF INVENTION: thereof and uses thereof, in particular for the diagnosis, prevention
; TITLE OF INVENTION: and treatment of infection
; FILE REFERENCE: 9710-003-999
; CURRENT APPLICATION NUMBER: US/09/198,452A
; CURRENT FILING DATE: 1998-11-24
; NUMBER OF SEQ ID NOS: 6849
; SEQ ID NO 849
; LENGTH: 228
; TYPE: PRT
; ORGANISM: Chlamydia pneumoniae
US-09-198-452A-849

Query Match          2.9%; Score 7; DB 4; Length 228;
Best Local Similarity 100.0%; Pred. No. 63;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      143 KDLGTLG 149
      |||||
Db      158 KDLGTLG 164

RESULT 7
US-09-134-000C-3673
; Sequence 3673, Application US/09134000C
; Patent No. 6617156
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
; TITLE OF INVENTION: ENTEROCOCCUS FAECALIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 032796-032
; CURRENT APPLICATION NUMBER: US/09/134,000C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/055,778
; PRIOR FILING DATE: 1997-08-15
; NUMBER OF SEQ ID NOS: 6812
; SOFTWARE: Patentin version 3.1
; SEQ ID NO 3673
; LENGTH: 274
; TYPE: PRT
; ORGANISM: Enterococcus faecalis
US-09-134-000C-3673

Query Match          2.9%; Score 7; DB 4; Length 274;
Best Local Similarity 100.0%; Pred. No. 74;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      33 SGLASAP 39
```

Db 89 SGLASAP 95

RESULT 8

US-09-252-991A-18308

; Sequence 18308, Application US/09252991A

; Patent No. 6551795

; GENERAL INFORMATION:

; APPLICANT: Marc J. Rubenfield et al.

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS

; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS

; FILE REFERENCE: 107195.136

; CURRENT APPLICATION NUMBER: US/09/252.991A

; CURRENT FILING DATE: 1999-02-18

; PRIOR APPLICATION NUMBER: US 60/074,788

; PRIOR FILING DATE: 1998-02-18

; PRIOR APPLICATION NUMBER: US 60/094,190

; PRIOR FILING DATE: 1998-07-27

; NUMBER OF SEQ ID NOS: 33142

; SEQ ID NO 18308

; LENGTH: 341

; TYPE: PRT

; ORGANISM: Pseudomonas aeruginosa

US-09-252-991A-18308

Query Match 2.9%; Score 7; DB 4; Length 341;

Best Local Similarity 100.0%; Pred. No. 91;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 117 PARSEAA 123

Db 284 PARSEAA 290

RESULT 9

US-09-489-039A-8827

; Sequence 8827, Application US/09489039A

; Patent No. 6610836

; GENERAL INFORMATION:

; APPLICANT: Gary Breton et. al

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA

; TITLE OF INVENTION: PNEUMONIAE FOR DIAGNOSTICS AND THERAPEUTICS

; FILE REFERENCE: 2709.2004001

; CURRENT APPLICATION NUMBER: US/09/489,039A

; CURRENT FILING DATE: 2000-01-27

; PRIOR APPLICATION NUMBER: US 60/117,747

; PRIOR FILING DATE: 1999-01-29

; NUMBER OF SEQ ID NOS: 14342

; SEQ ID NO 8827

; LENGTH: 348

; TYPE: PRT

; ORGANISM: Klebsiella pneumoniae

US-09-489-039A-8827

Query Match 2.9%; Score 7; DB 4; Length 348;

Best Local Similarity 100.0%; Pred. No. 93;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 126 QPVIGIS 132

Db 33 QPVIGIS 39

RESULT 10

US-09-107-532A-5053

; Sequence 5053, Application US/09107532A

; Patent No. 6583275

; GENERAL INFORMATION:

; APPLICANT: Lynn A. Doucette-Stamm and David Bush

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO

; TITLE OF INVENTION: ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS

; NUMBER OF SEQUENCES: 7310

US-09-107-532A-5053

Query Match 2.9%; Score 7; DB 4; Length 353;

Best Local Similarity 100.0%; Pred. No. 94;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 160 IIAIGAG 166

Db 207 IIAIGAG 213

RESULT 11

US-09-107-532A-4004

; Sequence 4004, Application US/09107532A

; Patent No. 6583275

; GENERAL INFORMATION:

; APPLICANT: Lynn A. Doucette-Stamm and David Bush

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO

; TITLE OF INVENTION: ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS

; NUMBER OF SEQUENCES: 7310

US-09-107-532A-4004

Query Match 2.9%; Score 7; DB 4; Length 353;

Best Local Similarity 100.0%; Pred. No. 94;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 160 IIAIGAG 166

Db 207 IIAIGAG 213

RESULT 12

US-09-107-532A-4004

; Sequence 4004, Application US/09107532A

; Patent No. 6583275

; GENERAL INFORMATION:

; APPLICANT: Lynn A. Doucette-Stamm and David Bush

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO

; TITLE OF INVENTION: ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS

; NUMBER OF SEQUENCES: 7310

US-09-107-532A-4004

Query Match 2.9%; Score 7; DB 4; Length 353;

Best Local Similarity 100.0%; Pred. No. 94;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 160 IIAIGAG 166

Db 207 IIAIGAG 213

RESULT 13

US-09-107-532A-4004

; Sequence 4004, Application US/09107532A

; Patent No. 6583275

; GENERAL INFORMATION:

; APPLICANT: Lynn A. Doucette-Stamm and David Bush

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO

; TITLE OF INVENTION: ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS

; NUMBER OF SEQUENCES: 7310

US-09-107-532A-4004

Query Match 2.9%; Score 7; DB 4; Length 353;

Best Local Similarity 100.0%; Pred. No. 94;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 160 IIAIGAG 166

Db 207 IIAIGAG 213

RESULT 14

US-09-107-532A-4004

; Sequence 4004, Application US/09107532A

; Patent No. 6583275

; GENERAL INFORMATION:

; APPLICANT: Lynn A. Doucette-Stamm and David Bush

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO

; TITLE OF INVENTION: ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS

; NUMBER OF SEQUENCES: 7310

US-09-107-532A-4004

Query Match 2.9%; Score 7; DB 4; Length 353;

Best Local Similarity 100.0%; Pred. No. 94;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 160 IIAIGAG 166

Db 207 IIAIGAG 213

RESULT 15

US-09-107-532A-4004

; Sequence 4004, Application US/09107532A

; Patent No. 6583275

; GENERAL INFORMATION:

; APPLICANT: Lynn A. Doucette-Stamm and David Bush

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO

; TITLE OF INVENTION: ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS

; NUMBER OF SEQUENCES: 7310

US-09-107-532A-4004

Query Match 2.9%; Score 7; DB 4; Length 353;

Best Local Similarity 100.0%; Pred. No. 94;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 160 IIAIGAG 166

Db 207 IIAIGAG 213

RESULT 16

US-09-107-532A-4004

; Sequence 4004, Application US/09107532A

; Patent No. 6583275

; GENERAL INFORMATION:

; APPLICANT: Lynn A. Doucette-Stamm and David Bush

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO

; TITLE OF INVENTION: ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS

; NUMBER OF SEQUENCES: 7310

US-09-107-532A-4004

Query Match 2.9%; Score 7; DB 4; Length 353;

Best Local Similarity 100.0%; Pred. No. 94;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 160 IIAIGAG 166

Db 207 IIAIGAG 213

RESULT 17

US-09-107-532A-4004

; Sequence 4004, Application US/09107532A

; Patent No. 6583275

; GENERAL INFORMATION:

; APPLICANT: Lynn A. Doucette-Stamm and David Bush

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO

; TITLE OF INVENTION: ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS

; NUMBER OF SEQUENCES: 7310

US-09-107-532A-4004

Query Match 2.9%; Score 7; DB 4; Length 353;

Best Local Similarity 100.0%; Pred. No. 94;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 160 IIAIGAG 166

Db 207 IIAIGAG 213

RESULT 18

US-09-107-532A-4004

; Sequence 4004, Application US/09107532A

; Patent No. 6583275

; GENERAL INFORMATION:

; APPLICANT: Lynn A. Doucette-Stamm and David Bush

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO

; TITLE OF INVENTION: ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS

; NUMBER OF SEQUENCES: 7310

US-09-107-532A-4004

Query Match 2.9%; Score 7; DB 4; Length 353;

Best Local Similarity 100.0%; Pred. No. 94;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 160 IIAIGAG 166

Db 207 IIAIGAG 213

RESULT 19

US-09-107-532A-4004

; Sequence 4004, Application US/09107532A

; Patent No. 6583275

; GENERAL INFORMATION:

; APPLICANT: Lynn A. Doucette-Stamm and David Bush

; TITLE OF INVENTION: NUCLE

APPLICATION NUMBER: US/09/107,532A  
FILING DATE: 30-Jun-1998  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 60/085,598  
FILING DATE: 14 May 1998  
APPLICATION NUMBER: 60/051571  
FILING DATE: July 2, 1997  
ATTORNEY/AGENT INFORMATION:  
NAME: Ariniello, Pamela Deneke  
REGISTRATION NUMBER: 40,489  
REFERENCE/DOCKET NUMBER: GTC-012  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (781)893-5007  
TELEFAX: (781)893-8277  
INFORMATION FOR SEQ ID NO: 4004:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 356 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
HYPOTHETICAL: YES  
ORIGINAL SOURCE:  
ORGANISM: Enterococcus faecium  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: (B) LOCATION 1...356  
SEQUENCE DESCRIPTION: SEQ ID NO: 4004:  
US-09-107-532A-4004

Query Match 2.9%; Score 7; DB 4; Length 356;  
Best Local Similarity 100.0%; Pred. No. 95;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 160 IIAIGAG 166  
|||||  
Db 212 IIAIGAG 218  
RESULT 12  
US-09-107-532A-6936  
Sequence 6936, Application US/09107532A  
Patent No. 6583275  
GENERAL INFORMATION:  
APPLICANT: Lynn A Doucette-Stamm and David Rush  
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: GENOME THERAPEUTICS CORPORATION  
STREET: 100 Beaver Street  
CITY: Waltham  
STATE: Massachusetts  
COUNTRY: USA  
ZIP: 02354  
COMPUTER READABLE FORM:  
MEDIUM TYPE: CD-ROM ISO9660  
COMPUTER: PC  
OPERATING SYSTEM: <Unknown>  
SOFTWARE: ASCII  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/107,532A  
FILING DATE: 30-Jun-1998  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 60/085,598  
FILING DATE: 14 May 1998  
APPLICATION NUMBER: 60/051571  
FILING DATE: July 2, 1997  
ATTORNEY/AGENT INFORMATION:  
NAME: Ariniello, Pamela Deneke  
REGISTRATION NUMBER: 40,489  
REFERENCE/DOCKET NUMBER: GTC-012  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (781)893-5007

TELEFAX: (781)893-8277  
INFORMATION FOR SEQ ID NO: 6936:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 356 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
HYPOTHETICAL: YES  
ORIGINAL SOURCE:  
ORGANISM: Enterococcus faecium  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: (B) LOCATION 1...356  
SEQUENCE DESCRIPTION: SEQ ID NO: 6936:  
US-09-107-532A-6936

Query Match 2.9%; Score 7; DB 4; Length 356;  
Best Local Similarity 100.0%; Pred. No. 95;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 166 GILGYS 172  
|||||  
Db 144 GILGYS 150

RESULT 13  
US-09-252-991A-29970  
Sequence 29970, Application US/09252991A  
Patent No. 6551795  
GENERAL INFORMATION:  
APPLICANT: Marc J. Rubenfield et al.  
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS  
FILE REFERENCE: 107196.136  
CURRENT APPLICATION NUMBER: US/09/252,991A  
CURRENT FILING DATE: 1999-02-18  
PRIOR APPLICATION NUMBER: US 60/074,788  
PRIOR FILING DATE: 1998-02-18  
PRIOR APPLICATION NUMBER: US 60/094,190  
PRIOR FILING DATE: 1998-07-27  
NUMBER OF SEQ ID NOS: 33142  
SEQ ID NO 29970  
LENGTH: 381  
TYPE: PRT  
ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-29970

Query Match 2.9%; Score 7; DB 4; Length 381;  
Best Local Similarity 100.0%; Pred. No. 1e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 50 RNPDEDP 56  
|||||  
Db 127 RNPDEDP 133

RESULT 14  
US-09-252-991A-28999  
Sequence 28999, Application US/09252991A  
Patent No. 6551795  
GENERAL INFORMATION:  
APPLICANT: Marc J. Rubenfield et al.  
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS  
FILE REFERENCE: 107196.136  
CURRENT APPLICATION NUMBER: US/09/252,991A  
CURRENT FILING DATE: 1999-02-18  
PRIOR APPLICATION NUMBER: US 60/074,788  
PRIOR FILING DATE: 1998-02-18  
PRIOR APPLICATION NUMBER: US 60/094,190  
PRIOR FILING DATE: 1998-07-27  
NUMBER OF SEQ ID NOS: 33142  
SEQ ID NO 28999

```
; LENGTH: 413
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-28999

Query Match      2.9%; Score 7; DB 4; Length 413;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      114 NALPARS 120
Db      184 NALPARS 190

RESULT 15
US-09-489-039A-13479
; Sequence 13479, Application US/09489039A
; Patent No. 6610836
; GENERAL INFORMATION:
; APPLICANT: Gary Breton et. al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA
; FILE REFERENCE: 2709.2004001
; CURRENT APPLICATION NUMBER: US/09/489,039A
; CURRENT FILING DATE: 2000-01-27
; PRIOR APPLICATION NUMBER: US 60/117,747
; PRIOR FILING DATE: 1999-01-29
; NUMBER OF SEQ ID NOS: 14342
; SEQ ID NO 13479
; LENGTH: 472
; TYPE: PRT
; ORGANISM: Klebsiella pneumoniae
US-09-489-039A-13479

Query Match      2.9%; Score 7; DB 4; Length 472;
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      18 SPAPGLR 24
Db      190 SPAPGLR 196

RESULT 16
US-08-009-075-4
; Sequence 4, Application US/08009075
; Patent No. 5300436
; GENERAL INFORMATION:
; APPLICANT: GOLDSTEIN, Menek
; APPLICANT: WU, Jing
; APPLICANT: FILER, David
; APPLICANT: FRIEDHOFF, Arnold J.
; TITLE OF INVENTION: GENETICALLY MODIFIED TYROSINE
; TITLE OF INVENTION: HYDROXYLASE AND USES THEREOF
; NUMBER OF SEQUENCES: 13
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: BROWDY and NEIMARK
; STREET: 419 Seventh Street, N.W., Suite 300
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20004
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/009,075
; FILING DATE: 19930126
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: TOWNSEND, GUY K.

; REGISTRATION NUMBER: 34,033
; REFERENCE/DOCKET NUMBER: GOLDSTEIN=LA
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-628-5197
; TELEFAX: 202-737-3528
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 497 amino acids
; TYPE: AMINO ACID
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-009-075-4

Query Match      2.9%; Score 7; DB 1; Length 497;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      121 EAAAVQP 127
Db      415 EAAAVQP 421

RESULT 17
US-09-489-039A-10153
; Sequence 10153, Application US/09489039A
; Patent No. 6610836
; GENERAL INFORMATION:
; APPLICANT: Gary Breton et. al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA
; FILE REFERENCE: 2709.2004001
; CURRENT APPLICATION NUMBER: US/09/489,039A
; CURRENT FILING DATE: 2000-01-27
; PRIOR APPLICATION NUMBER: US 60/117,747
; PRIOR FILING DATE: 1999-01-29
; NUMBER OF SEQ ID NOS: 14342
; SEQ ID NO 10153
; LENGTH: 517
; TYPE: PRT
; ORGANISM: Klebsiella pneumoniae
US-09-489-039A-10153

Query Match      2.9%; Score 7; DB 4; Length 517;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      34 GLASAPV 40
Db      312 GLASAPV 318

RESULT 18
US-08-467-822-27
; Sequence 27, Application US/08467822
; Patent No. 5843460
; GENERAL INFORMATION:
; APPLICANT: Labigne, Agnes
; APPLICANT: Sauerbaum, Sebastien
; APPLICANT: Ferrero, Richard L.
; APPLICANT: Thiberge, Jean-Michel
; TITLE OF INVENTION: IMMUNOGENIC COMPOSITIONS AGAINST
; TITLE OF INVENTION: HELICOBACTER INFECTION, POLYPEPTIDES FOR USE IN THE
; TITLE OF INVENTION: COMPOSITIONS, AND NUCLEIC ACID SEQUENCES ENCODING SAID
; TITLE OF INVENTION: POLYPEPTIDES
; NUMBER OF SEQUENCES: 44
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Finnegan, Henderson, Farabow, Garrett &
; ADDRESSEE: Dunner
; STREET: 1300 I Street, N.W.
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20005-3315
```

```
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/467,822
FILING DATE: 06-JUN-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/447,177
FILING DATE: 19-MAY-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/432,697
FILING DATE: 02-MAY-1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Meyers, Kenneth J.
REGISTRATION NUMBER: 25,146
REFERENCE/DOCKET NUMBER: 03495.0137-02000
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 408-4000
TELEFAX: (202) 408-4400
INFORMATION FOR SEQ ID NO: 27:
SEQUENCE CHARACTERISTICS:
LENGTH: 569 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-467-822-27

Query Match          2.9%; Score 7; DB 2; Length 569;
Best Local Similarity 100.0%; Pred. No. 1.5e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      66 EAGVPEK 72
      |||||
Db      498 EAGVPEK 504

RESULT 19
US-08-432-697-27
; Sequence 27, Application US/08432697
; Patent No. 6248330
; GENERAL INFORMATION:
; APPLICANT: Labigne, Agnes
; APPLICANT: Sauerbaum, Sebastien
; APPLICANT: Ferrero, Richard L.
; APPLICANT: Thiberge, Jean-Michel
; TITLE OF INVENTION: IMMUNOGENIC COMPOSITIONS AGAINST
; TITLE OF INVENTION: HELICOBLACTER INFECTION, POLYPEPTIDES FOR USE IN THE
; TITLE OF INVENTION: COMPOSITIONS, AND NUCLEIC ACID SEQUENCES ENCODING SAID
; NUMBER OF SEQUENCES: 44
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Finnegan, Henderson, Farabow, Garrett &
; STREET: 1300 I Street, N.W.
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20005-3315
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/432,697
; FILING DATE: 02-MAY-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Meyers, Kenneth J.
; REGISTRATION NUMBER: 25,146
; REFERENCE/DOCKET NUMBER: 03495.0137-02000
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 408-4000
; TELEFAX: (202) 408-4400
; INFORMATION FOR SEQ ID NO: 27:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 569 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-467-822-27

Query Match          2.9%; Score 7; DB 2; Length 569;
Best Local Similarity 100.0%; Pred. No. 1.5e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      66 EAGVPEK 72
      |||||
Db      498 EAGVPEK 504

RESULT 20
US-08-466-248-27
; Sequence 27, Application US/08466248
; Patent No. 6258359
; GENERAL INFORMATION:
; APPLICANT: Labigne, Agnes
; APPLICANT: Sauerbaum, Sebastien
; APPLICANT: Ferrero, Richard L.
; APPLICANT: Thiberge, Jean-Michel
; TITLE OF INVENTION: IMMUNOGENIC COMPOSITIONS AGAINST
; TITLE OF INVENTION: HELICOBLACTER INFECTION, POLYPEPTIDES FOR USE IN THE
; TITLE OF INVENTION: COMPOSITIONS, AND NUCLEIC ACID SEQUENCES ENCODING SAID
; NUMBER OF SEQUENCES: 44
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Finnegan, Henderson, Farabow, Garrett &
; STREET: 1300 I Street, N.W.
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20005-3315
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/466,248
; FILING DATE: 06-JUN-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/447,177
; FILING DATE: 19-MAY-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/432,697
; FILING DATE: 02-MAY-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Meyers, Kenneth J.
; REGISTRATION NUMBER: 25,146
; REFERENCE/DOCKET NUMBER: 03495.0137-02000
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 408-4000
; TELEFAX: (202) 408-4400
; INFORMATION FOR SEQ ID NO: 27:
; SEQUENCE CHARACTERISTICS:
```

LENGTH: 569 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-466-248-27

Query Match 2.9%; Score 7; DB 3; Length 569;  
Best Local Similarity 100.0%; Pred. No. 1.5e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 66 EAGVPEK 72  
DB 498 EAGVPEK 504

## RESULT 21

US-09-543-681A-6029  
Sequence 6029, Application US/09543681A

Patent No. 6605709

GENERAL INFORMATION:

APPLICANT: GARY BRETON

TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PROTEUS MIRABILIS

FILE REFERENCE: 2709, 1002-001

CURRENT APPLICATION NUMBER: US/09/543, 681A

CURRENT FILING DATE: 2000-04-05

PRIOR APPLICATION NUMBER: US 60/128,706

PRIOR FILING DATE: 1999-04-09

NUMBER OF SEQ ID NOS: 8344

SEQ ID NO 6029

LENGTH: 569

TYPE: PRT

ORGANISM: Proteus mirabilis

US-09-543-681A-6029

Query Match 2.9%; Score 7; DB 4; Length 569;  
Best Local Similarity 100.0%; Pred. No. 1.5e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 66 EAGVPEK 72  
DB 500 EAGVPEK 506

## RESULT 22

US-08-532-547-7

Sequence 7, Application US/08532547

Patent No. 5861543

GENERAL INFORMATION:

APPLICANT: LAMBERT, BART

APPLICANT: JANSSENS, STEFAN

APPLICANT: VAN AUDENHOVE, KATRIEN

APPLICANT: PEPEROEN, MARNIX

APPLICANT: VAN RIE, JEROEN

APPLICANT: VAN AARSSEN, ROEL

TITLE OF INVENTION: NEW BACILLUS THURINGIENSIS STRAINS AND THEIR

INSECTICIDAL PROTEINS.

NUMBER OF SEQUENCES: 9

CORRESPONDENCE ADDRESS:

ADDRESSEE: Birch, Stewart, Kolasch & Birch, LLP

STREET: P.O. Box 747

CITY: Falls Church

STATE: Virginia

COUNTRY: USA

ZIP: 22040-0747

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/532,547

FILING DATE: 06-DEC-1996  
CLASSIFICATION: 800  
ATTORNEY/AGENT INFORMATION:  
NAME: SVENSSON, LEONARD R.  
REGISTRATION NUMBER: 30,330  
REFERENCE/DOCKET NUMBER: 2121-109P  
TELEPHONE: (703) 205-8000  
TELEFAX: (703) 205-8050  
TELEX: 248345  
INFORMATION FOR SEQ ID NO: 7:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 625 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-532-547-7

Query Match 2.9%; Score 7; DB 2; Length 625;  
Best Local Similarity 100.0%; Pred. No. 1.6e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 36 ASAPVSG 42  
DB 482 ASAPVSG 488

## RESULT 23

US-08-532-547-9

Sequence 9, Application US/08532547

Patent No. 5861543

GENERAL INFORMATION:

APPLICANT: LAMBERT, BART

APPLICANT: JANSSENS, STEFAN

APPLICANT: VAN AUDENHOVE, KATRIEN

APPLICANT: PEPEROEN, MARNIX

APPLICANT: VAN RIE, JEROEN

APPLICANT: VAN AARSSEN, ROEL

TITLE OF INVENTION: NEW BACILLUS THURINGIENSIS STRAINS AND THEIR

INSECTICIDAL PROTEINS.

NUMBER OF SEQUENCES: 9

CORRESPONDENCE ADDRESS:

ADDRESSEE: Birch, Stewart, Kolasch & Birch, LLP

STREET: P.O. Box 747

CITY: Falls Church

STATE: Virginia

COUNTRY: USA

ZIP: 22040-0747

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/532,547

FILING DATE: 06-DEC-1996

CLASSIFICATION: 800

ATTORNEY/AGENT INFORMATION:

NAME: SVENSSON, LEONARD R.

REGISTRATION NUMBER: 30,330

REFERENCE/DOCKET NUMBER: 2121-109P

TELEPHONE: (703) 205-8000

TELEFAX: (703) 205-8050

TELEX: 248345

INFORMATION FOR SEQ ID NO: 9:

SEQUENCE CHARACTERISTICS:

LENGTH: 625 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

US-08-532-547-9

Query Match 2.9%; Score 7; DB 2; Length 625;  
Best Local Similarity 100.0%; Pred. No. 1.6e+02; Indels 0; Gaps 0;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 36 ASAPVSG 42  
Db 482 ASAPVSG 488

## RESULT 24

US-09-019-809-7  
; Sequence 7, Application US/09019809  
; Patent No. 6143550  
; GENERAL INFORMATION:  
; APPLICANT: LAMBERT, BART  
; APPLICANT: JANSSENS, STEFAN  
; APPLICANT: VAN AUDENHOVE, KATRIEN  
; APPLICANT: PEPEROEN, MARNIX  
; APPLICANT: VAN RIE, JEROEN  
; APPLICANT: VAN AARSEN, ROEL  
; TITLE OF INVENTION: NEW BACILLUS THURINGIENSIS STRAINS AND THEIR  
; TITLE OF INVENTION: INSECTICIDAL PROTEINS.  
; NUMBER OF SEQUENCES: 9  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Birch, Stewart, Kolasch & Birch, LLP  
; STREET: P.O. Box 747  
; CITY: Falls Church  
; STATE: Virginia  
; COUNTRY: USA  
; ZIP: 22040-0747  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25 (BPO)  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/019,809  
; FILING DATE: 02-FEB-1998  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: SVENSSON, LEONARD R.  
; REGISTRATION NUMBER: 30,330  
; REFERENCE/DOCKET NUMBER: 2121-135P  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (703) 205-8000  
; TELEFAX: (703) 205-8050  
; TELEX: 248345  
; INFORMATION FOR SEQ ID NO: 7:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 625 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein

US-09-019-809-7  
; Sequence 9, Application US/09019809  
; Patent No. 6143550  
; GENERAL INFORMATION:  
; APPLICANT: LAMBERT, BART  
; APPLICANT: JANSSENS, STEFAN  
; APPLICANT: VAN AUDENHOVE, KATRIEN  
; APPLICANT: PEPEROEN, MARNIX  
; APPLICANT: VAN RIE, JEROEN  
; APPLICANT: VAN AARSEN, ROEL  
; TITLE OF INVENTION: NEW BACILLUS THURINGIENSIS STRAINS AND THEIR  
; TITLE OF INVENTION: INSECTICIDAL PROTEINS.  
; NUMBER OF SEQUENCES: 9  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Birch, Stewart, Kolasch & Birch, LLP  
; STREET: P.O. Box 747  
; CITY: Falls Church  
; STATE: Virginia  
; COUNTRY: USA  
; ZIP: 22040-0747  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25 (BPO)  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/019,809  
; FILING DATE: 02-FEB-1998  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: SVENSSON, LEONARD R.  
; REGISTRATION NUMBER: 30,330  
; REFERENCE/DOCKET NUMBER: 2121-135P  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (703) 205-8000  
; TELEFAX: (703) 205-8050  
; TELEX: 248345  
; INFORMATION FOR SEQ ID NO: 7:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 625 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein

Query Match 2.9%; Score 7; DB 3; Length 625;  
Best Local Similarity 100.0%; Pred. No. 1.6e+02; Indels 0; Gaps 0;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 36 ASAPVSG 42  
Db 482 ASAPVSG 488

## RESULT 25

US-09-019-809-9  
; Sequence 9, Application US/09019809  
; Patent No. 6143550  
; GENERAL INFORMATION:  
; APPLICANT: LAMBERT, BART  
; APPLICANT: JANSSENS, STEFAN  
; APPLICANT: VAN AUDENHOVE, KATRIEN  
; APPLICANT: PEPEROEN, MARNIX

APPLICANT: VAN RIE, JEROEN  
APPLICANT: VAN AARSEN, ROEL  
TITLE OF INVENTION: NEW BACILLUS THURINGIENSIS STRAINS AND THEIR  
TITLE OF INVENTION: INSECTICIDAL PROTEINS.  
NUMBER OF SEQUENCES: 9  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Birch, Stewart, Kolasch & Birch, LLP  
STREET: P.O. Box 747  
CITY: Falls Church  
STATE: Virginia  
COUNTRY: USA  
ZIP: 22040-0747  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)  
CURRENT APPLICATION DATA: US/09/019,809  
FILING DATE: 02-FEB-1998  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: SVENSSON, LEONARD R.  
REGISTRATION NUMBER: 30,330  
REFERENCE/DOCKET NUMBER: 2121-135P  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703) 205-8000  
TELEFAX: (703) 205-8050  
TELEX: 248345  
INFORMATION FOR SEQ ID NO: 9:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 625 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein

US-09-019-809-9  
; Sequence 7, Application US/09471177  
; Patent No. 6448226  
; GENERAL INFORMATION:  
; APPLICANT: LAMBERT, BART  
; APPLICANT: JANSSENS, STEFAN  
; APPLICANT: VAN AUDENHOVE, KATRIEN  
; APPLICANT: PEPEROEN, MARNIX  
; APPLICANT: VAN RIE, JEROEN  
; APPLICANT: VAN AARSEN, ROEL  
; TITLE OF INVENTION: NEW BACILLUS THURINGIENSIS STRAINS AND THEIR  
; TITLE OF INVENTION: INSECTICIDAL PROTEINS.  
; NUMBER OF SEQUENCES: 9  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Birch, Stewart, Kolasch & Birch, LLP  
; STREET: P.O. Box 747  
; CITY: Falls Church  
; STATE: Virginia  
; COUNTRY: USA  
; ZIP: 22040-0747  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)  
; CURRENT APPLICATION DATA:

Query Match 2.9%; Score 7; DB 3; Length 625;  
Best Local Similarity 100.0%; Pred. No. 1.6e+02; Indels 0; Gaps 0;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 36 ASAPVSG 42  
Db 482 ASAPVSG 488

## RESULT 26

US-09-471-177-7  
; Sequence 7, Application US/09471177  
; Patent No. 6448226  
; GENERAL INFORMATION:  
; APPLICANT: LAMBERT, BART  
; APPLICANT: JANSSENS, STEFAN  
; APPLICANT: VAN AUDENHOVE, KATRIEN  
; APPLICANT: PEPEROEN, MARNIX  
; APPLICANT: VAN RIE, JEROEN  
; APPLICANT: VAN AARSEN, ROEL  
; TITLE OF INVENTION: NEW BACILLUS THURINGIENSIS STRAINS AND THEIR  
; TITLE OF INVENTION: INSECTICIDAL PROTEINS.  
; NUMBER OF SEQUENCES: 9  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Birch, Stewart, Kolasch & Birch, LLP  
; STREET: P.O. Box 747  
; CITY: Falls Church  
; STATE: Virginia  
; COUNTRY: USA  
; ZIP: 22040-0747  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)  
; CURRENT APPLICATION DATA:

US-09-471-177-7  
; Sequence 7, Application US/09471177  
; Patent No. 6448226  
; GENERAL INFORMATION:  
; APPLICANT: LAMBERT, BART  
; APPLICANT: JANSSENS, STEFAN  
; APPLICANT: VAN AUDENHOVE, KATRIEN  
; APPLICANT: PEPEROEN, MARNIX  
; APPLICANT: VAN RIE, JEROEN  
; APPLICANT: VAN AARSEN, ROEL  
; TITLE OF INVENTION: NEW BACILLUS THURINGIENSIS STRAINS AND THEIR  
; TITLE OF INVENTION: INSECTICIDAL PROTEINS.  
; NUMBER OF SEQUENCES: 9  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Birch, Stewart, Kolasch & Birch, LLP  
; STREET: P.O. Box 747  
; CITY: Falls Church  
; STATE: Virginia  
; COUNTRY: USA  
; ZIP: 22040-0747  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)  
; CURRENT APPLICATION DATA:

US-09-019-809-7  
; Sequence 9, Application US/09019809  
; Patent No. 6143550  
; GENERAL INFORMATION:  
; APPLICANT: LAMBERT, BART  
; APPLICANT: JANSSENS, STEFAN  
; APPLICANT: VAN AUDENHOVE, KATRIEN  
; APPLICANT: PEPEROEN, MARNIX

APPLICATION NUMBER: US/09/471,177  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/09/019,809  
FILING DATE: 02-FEB-1998  
ATTORNEY/AGENT INFORMATION:  
NAME: SVENSSON, LEONARD R.  
REGISTRATION NUMBER: 30,330  
REFERENCE/DOCKET NUMBER: 2121-135P  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703) 205-8000  
TELEFAX: (703) 205-8050  
TELEX: 248345  
INFORMATION FOR SEQ ID NO: 7:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 625 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-09-471-177-7

Query Match 2.9%; Score 7; DB 4; Length 625;  
Best Local Similarity 100.0%; Pred. No. 1.6e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 36 ASAPVSG 42  
|||||  
Db 482 ASAPVSG 488

## RESULT 27

US-09-471-177-9  
Sequence 9, Application US/09471177  
Patent No. 6448226  
GENERAL INFORMATION:  
APPLICANT: LAMBERT, BART  
APPLICANT: JANSSENS, STEFAN  
APPLICANT: VAN AUDENHOVE, KATHRIEN  
APPLICANT: PEERDEN, MARINX  
APPLICANT: VAN RIE, JEROEN  
APPLICANT: VAN AARSEN, ROEL  
TITLE OF INVENTION: NEW BACILLUS THURINGIENSIS STRAINS AND THEIR  
TITLE OF INVENTION: INSECTICIDAL PROTEINS.  
NUMBER OF SEQUENCES: 9  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Birch, Stewart, Kolasch & Birch, LLP  
STREET: P.O. Box 747  
CITY: Falls Church  
STATE: Virginia  
COUNTRY: USA  
ZIP: 22040-0747  
COMPUTER READABLE FORM: disk  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.25 (EPO)  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/471,177  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/09/019,809  
FILING DATE: 02-FEB-1998  
ATTORNEY/AGENT INFORMATION:  
NAME: SVENSSON, LEONARD R.  
REGISTRATION NUMBER: 30,330  
REFERENCE/DOCKET NUMBER: 2121-135P  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703) 205-8000  
TELEFAX: (703) 205-8050  
TELEX: 248345  
INFORMATION FOR SEQ ID NO: 9:

SEQUENCE CHARACTERISTICS:  
LENGTH: 625 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-09-471-177-9

Query Match 2.9%; Score 7; DB 4; Length 625;  
Best Local Similarity 100.0%; Pred. No. 1.6e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 36 ASAPVSG 42  
|||||  
Db 482 ASAPVSG 488

## RESULT 28

US-09-946-678-2  
Sequence 2, Application US/09946678  
Patent No. 6541236  
GENERAL INFORMATION:  
APPLICANT: ITO, Kotaro  
APPLICANT: UMITSUKI, Genryou  
APPLICANT: KOYAMA, Yasuji  
TITLE OF INVENTION: Protein Having Glutaminase Activity and Gene Encoding the Same  
FILE REFERENCE: 0283-0158P  
CURRENT APPLICATION NUMBER: US/09/946,678  
CURRENT FILING DATE: 2001-09-06  
PRIOR APPLICATION NUMBER: JP 2000-270371  
PRIOR FILING DATE: 2000-09-06  
NUMBER OF SEQ ID NOS: 27  
SOFTWARE: Patent in version 3.1  
SEQ ID NO 2  
LENGTH: 684  
TYPE: PRT  
ORGANISM: Cryptococcus nodaensis  
US-09-946-678-2

Query Match 2.9%; Score 7; DB 4; Length 684;  
Best Local Similarity 100.0%; Pred. No. 1.7e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 30 DAQSGLA 36  
|||||  
Db 411 DAQSGLA 417

## RESULT 29

US-08-469-537A-72  
Sequence 72, Application US/08469537A  
Patent No. 5843749  
GENERAL INFORMATION:  
APPLICANT: Maisompierre, et al.  
TITLE OF INVENTION: EHK AND ROR TYROSINE  
TITLE OF INVENTION: KINASES  
NUMBER OF SEQUENCES: 107  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Regeneron Pharmaceuticals, Inc.  
STREET: 777 Old Saw Mill River Road  
CITY: Tarrytown  
STATE: NY  
COUNTRY: U.S.A.  
ZIP: 10591  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FastSeq Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/469,537A  
FILING DATE: 06-JUN-1995  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:



APPLICATION NUMBER: USSN 08/406,247  
FILING DATE: 17-MAR-1995  
APPLICATION NUMBER: USSN 08/144,992  
FILING DATE: 28-OCT-1993  
APPLICATION NUMBER: USSN 07/736,559  
FILING DATE: 26-JUL-1991  
ATTORNEY/AGENT INFORMATION:  
NAME: Kempler, Ph.D., Gail M  
REGISTRATION NUMBER: 32,143  
REFERENCE/DOCKET NUMBER: REG 070C  
TELEPHONE: 914-345-7400  
TELEFAX: 914-345-7721  
TELEX:  
INFORMATION FOR SEQ ID NO: 72:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 800 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: unknown  
MOLECULE TYPE: peptide  
US-08-469-537A-72

Query Match 2.9%; Score 7; DB 2; Length 800;  
Best Local Similarity 100.0%; Pred. No. 2e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 46 HSYCRNP 52  
DB 222 HSYCRNP 228

RESULT 30  
US-08-469-537A-78  
Sequence 78, Application US/08469537A  
Patent No. 5843749  
GENERAL INFORMATION:  
APPLICANT: Maisonnier, et al.  
TITLE OF INVENTION: EHK AND ROR TYROSINE  
NUMBER OF SEQUENCES: 107  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Regeneron Pharmaceuticals, Inc.  
STREET: 777 Old Saw Mill River Road  
CITY: Tarrytown  
STATE: NY  
COUNTRY: U.S.A.  
ZIP: 10591  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FastSeq Version 2.0  
CURRENT APPLICATION DATA:  
FILING DATE: 06-JUN-1995  
CLASSIFICATION: 435  
PRIOR APPLICATION NUMBER: USSN 08/406,247  
FILING DATE: 17-MAR-1995  
APPLICATION NUMBER: USSN 08/144,992  
FILING DATE: 28-OCT-1993  
APPLICATION NUMBER: USSN 07/736,559  
FILING DATE: 26-JUL-1991  
ATTORNEY/AGENT INFORMATION:  
NAME: Kempler, Ph.D., Gail M  
REGISTRATION NUMBER: 32,143  
REFERENCE/DOCKET NUMBER: REG 070C  
TELEPHONE: 914-345-7400  
TELEFAX: 914-345-7721  
TELEX:  
INFORMATION FOR SEQ ID NO: 78:

SEQUENCE CHARACTERISTICS:  
LENGTH: 800 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-469-537A-78

Query Match 2.9%; Score 7; DB 2; Length 800;  
Best Local Similarity 100.0%; Pred. No. 2e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 46 HSYCRNP 52  
DB 222 HSYCRNP 228

RESULT 31  
US-09-269-861A-8  
Sequence 8, Application US/09269861A  
Patent No. 6468775  
GENERAL INFORMATION:  
APPLICANT: Ankenbauer, Waltraud  
APPLICANT: Markau, Ursula  
APPLICANT: Svetlichny, Vitaly  
APPLICANT: Schmitz-Agheguyan, Gudrun  
APPLICANT: Reiser, Astrid  
APPLICANT: Angerer, Bernhard  
APPLICANT: Ebenbichler, Christine  
APPLICANT: Laue, Frank  
APPLICANT: Bonch-Osmolovskaya, Elizaveta  
TITLE OF INVENTION: THERMOSTABLE DNA POLYMERASE FROM CARBOXYDOTHERMUS HYDROGENOFORMI  
FILE REFERENCE: 4494  
CURRENT APPLICATION NUMBER: US/09/269,861A  
CURRENT FILING DATE: 1999-11-22  
PRIOR APPLICATION NUMBER: PCT/EP97/05391  
PRIOR FILING DATE: 1997-10-01  
PRIOR APPLICATION NUMBER: EP/96115873.0  
PRIOR FILING DATE: 1996-10-03  
NUMBER OF SEQ ID NOS: 9  
SOFTWARE: PatentIn version 3.0  
SEQ ID NO 8  
LENGTH: 831  
TYPE: PRT  
ORGANISM: Carboxydothermus hydrogenoformans  
US-09-269-861A-8

Query Match 2.9%; Score 7; DB 4; Length 831;  
Best Local Similarity 100.0%; Pred. No. 2.1e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 210 VDEKTVV 216  
DB 140 VDEKTVV 146

RESULT 32  
US-08-469-537A-105  
Sequence 105, Application US/08469537A  
Patent No. 5843749  
GENERAL INFORMATION:  
APPLICANT: Maisonnier, et al.  
TITLE OF INVENTION: EHK AND ROR TYROSINE  
NUMBER OF SEQUENCES: 107  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Regeneron Pharmaceuticals, Inc.  
STREET: 777 Old Saw Mill River Road  
CITY: Tarrytown  
STATE: NY  
COUNTRY: U.S.A.  
ZIP: 10591  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette

COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FastSeq Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/469,537A  
FILING DATE: 06-JUN-1995  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: USSN 08/406,247  
FILING DATE: 17-MAR-1995  
APPLICATION NUMBER: USSN 08/144,992  
FILING DATE: 28-OCT-1993  
APPLICATION NUMBER: USSN 07/736,559  
FILING DATE: 26-JUL-1991  
ATTORNEY/AGENT INFORMATION:  
NAME: Kempster, Ph.D., Gail M  
REGISTRATION NUMBER: 32,143  
REFERENCE/DOCKET NUMBER: REG 070C  
TELEPHONE: 914-345-7400  
TELEFAX: 914-345-7721  
TELEX:

## INFORMATION FOR SEQ ID NO: 105:

SEQUENCE CHARACTERISTICS:  
LENGTH: 937 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
FRAGMENT TYPE: internal  
FEATURE:

NAME/KEY: Human ROR1  
LOCATION: 1...937  
OTHER INFORMATION:  
US-08-469-537A-105

Query Match 2.9%; Score 7; DB 2; Length 937;  
Best Local Similarity 100.0%; Pred. No. 2.3e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 46 HSYCRNP 52  
Db 359 HSYCRNP 365

RESULT 33  
US-09-002-285-72  
Sequence 72, Application US/09002285  
Patent No. 636213  
GENERAL INFORMATION:  
APPLICANT: Schnepf, H. Ernest  
APPLICANT: Wicker, Carol  
APPLICANT: Narva, Kenneth E.  
APPLICANT: Walz, Michelle  
APPLICANT: Stockhoff, Brian  
APPLICANT: Muller-Cohn, Judy  
TITLE OF INVENTION: Toxins Active Against Pests  
NUMBER OF SEQUENCES: 105  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Saliwanchik, Lloyd & Saliwanchik  
STREET: 2421 N.W. 41st Street, Suite A-1  
CITY: Gainesville  
STATE: Florida  
COUNTRY: USA  
ZIP: 32606  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/002,285  
FILING DATE:

CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/886,615  
FILING DATE: 1-JUL-1997  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/674,002  
FILING DATE: 1-JUL-1996  
ATTORNEY/AGENT INFORMATION:  
NAME: Sanders, Jay M.  
REGISTRATION NUMBER: 39,355  
REFERENCE/DOCKET NUMBER: MA-701C2  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (352) 375-8100  
TELEFAX: (352) 372-5800  
INFORMATION FOR SEQ ID NO: 72:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1156 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-09-002-285-72

Query Match 2.9%; Score 7; DB 4; Length 1156;  
Best Local Similarity 100.0%; Pred. No. 2.8e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 36 ASAPVSG 42  
Db 523 ASAPVSG 529

RESULT 34  
US-09-589-477-72  
Sequence 72, Application US/09589477  
Patent No. 6570005  
GENERAL INFORMATION:  
APPLICANT: Schnepf, H. Ernest  
APPLICANT: Wicker, Carol  
APPLICANT: Narva, Kenneth E.  
APPLICANT: Walz, Michelle  
APPLICANT: Stockhoff, Brian  
TITLE OF INVENTION: Toxins Active Against Pests  
NUMBER OF SEQUENCES: 105  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Saliwanchik, Lloyd & Saliwanchik  
STREET: 2421 N.W. 41st Street, Suite A-1  
CITY: Gainesville  
STATE: Florida  
COUNTRY: USA  
ZIP: 32606  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/589,477  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/886,615  
FILING DATE: 1-JUL-1997  
CLASSIFICATION:  
ATTORNEY/AGENT INFORMATION:  
NAME: Sanders, Jay M.  
REGISTRATION NUMBER: 39,355  
REFERENCE/DOCKET NUMBER: MA-701C1C1  
TELECOMMUNICATION INFORMATION:

```
;
; TELEPHONE: (352) 375-8100
; TELEFAX: (352) 372-5800
; INFORMATION FOR SEQ ID NO: 72:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1156 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-09-589-477-72

Query Match      2.9%; Score 7; DB 4; Length 1156;
Best Local Similarity 100.0%; Pred. No. 2.8e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      36 ASAPVSG 42
Db      523 ASAPVSG 529

RESULT 35
US-09-661-322A-28
; Sequence 28, Application US/09661322A
; Patent No. 6593293
; GENERAL INFORMATION:
; APPLICANT: Baum, James A.
; APPLICANT: Chu, Chih-Rei
; APPLICANT: Donovan, William P.
; APPLICANT: Gilmer, Amy J.
; APPLICANT: Rugar, Mark J.
; TITLE OF INVENTION: Lepidopteran-Active Bacillus thuringiensis Delta-Endotoxin Compo
; FILE REFERENCE: MECO201
; CURRENT APPLICATION NUMBER: US/09/661.322A
; CURRENT FILING DATE: 2000-09-13
; NUMBER OF SEQ ID NOS: 63
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 28
; LENGTH: 1156
; TYPE: PRT
; ORGANISM: Bacillus thuringiensis
; US-09-661-322A-28

Query Match      2.9%; Score 7; DB 4; Length 1156;
Best Local Similarity 100.0%; Pred. No. 2.8e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      36 ASAPVSG 42
Db      523 ASAPVSG 529

RESULT 36
US-08-532-547-5
; Sequence 5, Application US/08532547
; Patent No. 5861543
; GENERAL INFORMATION:
; APPLICANT: LAMBERT, BART
; APPLICANT: JANSENS, STEFAN
; APPLICANT: VAN AUDENHOVE, KATRIEN
; APPLICANT: VAN RIE, JEROEN
; APPLICANT: VAN AARSEN, ROEL
; TITLE OF INVENTION: NEW BACILLUS THURINGIENSIS STRAINS AND THEIR
; TITLE OF INVENTION: INSECTICIDAL PROTEINS.
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Birch, Stewart, Kolasch & Birch, LLP
; STREET: P.O. Box 747
; CITY: Falls Church
; STATE: Virginia
; COUNTRY: USA
; ZIP: 22040-0747

;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25 (BPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/532.547
; FILING DATE: 06-DEC-1996
; CLASSIFICATION: 800
; ATTORNEY/AGENT INFORMATION:
; NAME: SVENSSON, LEONARD R.
; REGISTRATION NUMBER: 30,330
; REFERENCE/DOCKET NUMBER: 2121-109P
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 205-8000
; TELEFAX: (703) 205-8050
; TELEX: 248345
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1157 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-532-547-5

Query Match      2.9%; Score 7; DB 2; Length 1157;
Best Local Similarity 100.0%; Pred. No. 2.8e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      36 ASAPVSG 42
Db      523 ASAPVSG 529

RESULT 37
US-08-379-656B-5
; Sequence 5, Application US/08379656B
; Patent No. 5885571
; GENERAL INFORMATION:
; APPLICANT: Lambert Bart
; APPLICANT: Janssens, Stefan
; APPLICANT: Van Audenhove, Katrien
; APPLICANT: Peferoen, Marnix
; TITLE OF INVENTION: NEW BACILLUS THURINGIENSIS STRAINS AND
; TITLE OF INVENTION: THEIR INSECTICIDAL PROTEINS
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: BIRCH, STEWART, KOLASCH & BIRCH
; STREET: 8110 Gatehouse Road, Suite 500 East
; CITY: Falls Church
; STATE: Virginia
; COUNTRY: U.S.A.
; ZIP: 22042
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/379.656B
; FILING DATE: 23-MAR-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/EP93/01820
; FILING DATE: 12-JULY-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: EP 93400949.9
; FILING DATE: 09-APR-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: EP 92402358.8
; FILING DATE: 27-AUG-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Svensson, Leonard R.
; REGISTRATION NUMBER: 30,330
```

REFERENCE/DOCKET NUMBER: 2121-104P  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703) 205-8000  
TELEFAX: (703) 205-8050  
TELEX: 248345  
INFORMATION FOR SEQ ID NO: 5:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1157 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-379-656B-5

Query Match 2.9%; Score 7; DB 2; Length 1157;  
Best Local Similarity 100.0%; Pred. No. 2.8e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 36 ASAPVSG 42  
DB 523 ASAPVSG 529

RESULT 38

US-08-455-838-5  
Sequence 5, Application US/08455838  
Patent No. 6028246

GENERAL INFORMATION:  
APPLICANT: Lambert Bart  
APPLICANT: Jansens, Stefan  
APPLICANT: Van Audenhove, Katrien  
APPLICANT: Peferoen, Marnix  
TITLE OF INVENTION: NEW BACILLUS THURINGIENSIS STRAINS AND  
THEIR INSECTICIDAL PROTEINS  
NUMBER OF SEQUENCES: 5  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: BIRCH, STEWART, KOLASCH & BIRCH  
STREET: 8110 Gatehouse Road, Suite 500 East  
CITY: Falls Church  
STATE: Virginia  
COUNTRY: U.S.A.  
ZIP: 22042

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent in Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/455,838  
FILING DATE: 31-MAY-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/EP93/01820  
FILING DATE: 12-JULY-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: EP 93400949.9  
FILING DATE: 09-APR-1993

PRIOR APPLICATION DATA:  
APPLICATION NUMBER: EP 92402358.8  
FILING DATE: 27-AUG-1992  
ATTORNEY/AGENT INFORMATION:  
NAME: Svensson, Leonard R.  
REGISTRATION NUMBER: 30,330  
REFERENCE/DOCKET NUMBER: 2121-106P  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703) 205-8000  
TELEFAX: (703) 205-8050  
TELEX: 248345  
INFORMATION FOR SEQ ID NO: 5:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1157 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-455-838-5

Query Match 2.9%; Score 7; DB 3; Length 1157;  
Best Local Similarity 100.0%; Pred. No. 2.8e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 36 ASAPVSG 42  
DB 523 ASAPVSG 529

RESULT 39

US-09-019-809-5  
Sequence 5, Application US/09019809  
Patent No. 6143550

GENERAL INFORMATION:  
APPLICANT: LAMBERT, BART  
APPLICANT: JANSSENS, STEFAN  
APPLICANT: VAN AUDENHOVE, KATRIEN  
APPLICANT: PEPEROEN, MARNIX  
APPLICANT: VAN RIE, JEROEN  
APPLICANT: VAN AARSSEN, ROEL  
TITLE OF INVENTION: NEW BACILLUS THURINGIENSIS STRAINS AND THEIR  
INSECTICIDAL PROTEINS.  
NUMBER OF SEQUENCES: 9  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Birch, Stewart, Kolasch & Birch, LLP  
STREET: P.O. Box 747  
CITY: Falls Church  
STATE: Virginia  
COUNTRY: USA  
ZIP: 22040-0747

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent in Release #1.0, Version #1.25 (EPO)  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/019,809  
FILING DATE: 02-FEB-1998  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: SVENSSON, LEONARD R.  
REGISTRATION NUMBER: 30,330  
REFERENCE/DOCKET NUMBER: 2121-135P  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703) 205-8000  
TELEFAX: (703) 205-8050  
TELEX: 248345

INFORMATION FOR SEQ ID NO: 5:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1157 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-09-019-809-5

Query Match 2.9%; Score 7; DB 3; Length 1157;  
Best Local Similarity 100.0%; Pred. No. 2.8e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 36 ASAPVSG 42  
DB 523 ASAPVSG 529

RESULT 40

US-09-471-177-5  
Sequence 5, Application US/09471177  
Patent No. 6448226

GENERAL INFORMATION:  
APPLICANT: LAMBERT, BART  
APPLICANT: JANSSENS, STEFAN  
APPLICANT: VAN AUDENHOVE, KATRIEN

APPLICANT: PEPEROEN, MARNIX  
APPLICANT: VAN RIE, JEROEN  
APPLICANT: VAN AARSEN, ROEL  
TITLE OF INVENTION: NEW BACILLUS THURINGIENSIS STRAINS AND THEIR  
TITLE OF INVENTION: INSECTICIDAL PROTEINS.  
NUMBER OF SEQUENCES: 9  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Birch, Stewart, Kolasch & Birch, LLP  
STREET: P.O. Box 747  
CITY: Falls Church  
STATE: Virginia  
COUNTRY: USA  
ZIP: 22040-0747  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.25 (BPO)  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/471,177  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/09/019,809  
FILING DATE: 02-FEB-1998  
ATTORNEY/AGENT INFORMATION:  
NAME: SVENSSON, LEONARD R.  
REGISTRATION NUMBER: 30,330  
REFERENCE/DOCKET NUMBER: 2121-135P  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703) 205-8000  
TELEFAX: (703) 205-8050  
TELEX: 248345  
INFORMATION FOR SEQ ID NO: 5:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1157 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-09-471-177-5

Query Match 2.9%; Score 7; DB 4; Length 1157;  
Best Local Similarity 100.0%; Pred. No. 2.8e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 36 ASAPVSG 42  
|||||  
DB 523 ASAPVSG 529

RESULT 41  
US-09-252-991A-32960  
Sequence 32960, Application US/09252991A  
Patent No. 6551795  
GENERAL INFORMATION:  
APPLICANT: Marc J. Rubenfield et al.  
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS  
FILE REFERENCE: 107196.136  
CURRENT APPLICATION NUMBER: US/09/252,991A  
CURRENT FILING DATE: 1999-02-18  
PRIOR APPLICATION NUMBER: US 60/074,788  
PRIOR FILING DATE: 1998-02-18  
PRIOR APPLICATION NUMBER: US 60/094,190  
PRIOR FILING DATE: 1998-07-27  
NUMBER OF SEQ ID NOS: 33142  
SEQ ID NO 32960  
LENGTH: 1247  
TYPE: PRT  
ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-32960

Query Match 2.9%; Score 7; DB 4; Length 1247;

APPLICANT: PAPEROFF, JACKIE  
APPLICANT: Megabios Corporation  
APPLICANT: Pfizer, Inc.  
TITLE OF INVENTION: Inhibition of Angiogenesis by Delivery of Nucleic Acids  
TITLE OF INVENTION: Encoding Anti-Angiogenesis Polypeptides  
FILE REFERENCE: 018484-000110US  
CURRENT APPLICATION NUMBER: US/09/192,012A  
CURRENT FILING DATE: 1998-11-13  
EARLIER APPLICATION NUMBER: US 60/066,020  
EARLIER FILING DATE: 1997-11-14  
NUMBER OF SEQ ID NOS: 9  
SOFTWARE: Patent In Ver. 2.0  
SEQ ID NO 8  
LENGTH: 7  
TYPE: PRT  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: Description of Artificial Sequence:consensus  
OTHER INFORMATION: kringle domain  
US-09-192-012-8

Query Match 2.5%; Score 6; DB 4; Length 7;  
Best Local Similarity 100.0%; Pred. No. 3e+05;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 48 YCRNPD 53  
|||||  
DB 2 YCRNPD 7

RESULT 43  
US-08-267-092A-12  
Sequence 12, Application US/08267092A  
Patent No. 5599686  
GENERAL INFORMATION:  
APPLICANT: DeFeo-Jones, Deborah  
APPLICANT: Garsky, Victor M.  
APPLICANT: Jones, Raymond E.  
APPLICANT: Oliff, Allen I.  
TITLE OF INVENTION: NOVEL PEPTIDES  
NUMBER OF SEQUENCES: 65  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: David A. Muthard  
STREET: 126 E. Lincoln Avenue, P.O. Box 2000  
CITY: Rahway  
STATE: New Jersey  
COUNTRY: USA  
ZIP: 07065-0900  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/267,092A  
FILING DATE:  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Muthard, David A.  
REGISTRATION NUMBER: 35,297

```

; REFERENCE/DOCKET NUMBER: 19253
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (908) 594-3903
; TELEFAX: (908) 594-4720
; INFORMATION FOR SEQ ID NO: 12:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 8 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; FRAGMENT TYPE: internal
;
US-08-267-092A-12
;
Query Match 2.5%; Score 6; DB 1; Length 8;
Best Local Similarity 100.0%; Pred. No. 3e+05;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 237 AGTPGA 242
Db 3 AGTPGA 8

RESULT 44
US-08-468-161-12
; Sequence 12, Application US/08468161
; Patent No. 6143864
; GENERAL INFORMATION:
; APPLICANT: Defeo-Jones, Deborah
; APPLICANT: Feng, Dong-Mei
; APPLICANT: Garsky, Victor M.
; APPLICANT: Jones, Raymond E.
; APPLICANT: Oliff, Allen I.
; TITLE OF INVENTION: NOVEL PEPTIDES
; NUMBER OF SEQUENCES: 146
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: DAVID A. MUTHARD
; STREET: 126 E. Lincoln Avenue, P.O. BOX 2000
; CITY: RAHWAY
; STATE: NEW JERSEY
; COUNTRY: U.S.A.
; ZIP: 07065
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08468,161
; FILING DATE:
; CLASSIFICATION: 530
; ATTORNEY/AGENT INFORMATION:
; NAME: Muthard, David A.
; REGISTRATION NUMBER: 35,297
; REFERENCE/DOCKET NUMBER: 192531B
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (908) 594-3903
; TELEFAX: (908) 594-4720
; INFORMATION FOR SEQ ID NO: 12:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 8 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; FRAGMENT TYPE: internal
;
PCT-US95-08156-12
;
Query Match 2.5%; Score 6; DB 5; Length 8;
Best Local Similarity 100.0%; Pred. No. 3e+05;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 237 AGTPGA 242
Db 3 AGTPGA 8

RESULT 45
PCT-US95-08156-12
; Sequence 12, Application PC/TUS9508156
; GENERAL INFORMATION:
; APPLICANT: Defeo-Jones, Deborah
; APPLICANT: Feng, Dong-Mei
; APPLICANT: Garsky, Victor M.
; APPLICANT: Jones, Raymond E.
; APPLICANT: Oliff, Allen I.
; TITLE OF INVENTION: NOVEL PEPTIDES
; NUMBER OF SEQUENCES: 146
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: DAVID A. MUTHARD
; STREET: 126 E. Lincoln Avenue, P.O. BOX 2000
; CITY: RAHWAY
; STATE: NEW JERSEY
; COUNTRY: U.S.A.
; ZIP: 07065
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/08156
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Muthard, David A.
; REGISTRATION NUMBER: 35,297
; REFERENCE/DOCKET NUMBER: 192531B
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (908) 594-3903
; TELEFAX: (908) 594-4720
; INFORMATION FOR SEQ ID NO: 12:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 8 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; FRAGMENT TYPE: internal
;
PCT-US95-08156-12
;
Query Match 2.5%; Score 6; DB 5; Length 8;
Best Local Similarity 100.0%; Pred. No. 3e+05;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 237 AGTPGA 242
Db 3 AGTPGA 8

RESULT 46
US-09-348-953-4
; Sequence 4, Application US/09348953
; Patent No. 6538103
; GENERAL INFORMATION:
; APPLICANT: Ji, Richard Wei-Dong
; APPLICANT: Trill, Pamela A.
; TITLE OF INVENTION: LYSINE BINDING FRAGMENTS OF ANGIOTATIN
; FILE REFERENCE: DB11sequences
; CURRENT APPLICATION NUMBER: US/09/348,953

```

```
; CURRENT FILING DATE: 1999-07-07
; PRIOR APPLICATION NUMBER: 60/092,831
; PRIOR FILING DATE: 1998-07-14
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 4
; LENGTH: 38
; TYPE: PRT
; ORGANISM: human
US-09-348-953-4

Query Match      2.5%; Score 6; DB 4; Length 38;
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      48 YCRNPD 53
      |||||
Db      16 YCRNPD 21

RESULT 47
US-09-227-357-589
; Sequence 589, Application US/09227357
; Patent No. 6342581
; GENERAL INFORMATION:
; APPLICANT: Fischer et al.
; TITLE OF INVENTION: 123 Human Secreted Proteins
; FILE REFERENCE: P2010P1
; CURRENT APPLICATION NUMBER: US/09/227,357
; CURRENT FILING DATE: 1999-01-08
; EARLIER APPLICATION NUMBER: PCT/US98/13684
; EARLIER FILING DATE: 1998-07-07
; EARLIER APPLICATION NUMBER: 60/051,926
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/052,793
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/051,925
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/051,929
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/052,803
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/052,732
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/051,931
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/051,932
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/051,916
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/051,930
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/051,918
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/051,920
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/052,733
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/052,795
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/051,919
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/051,928
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/055,722
; EARLIER FILING DATE: 1997-08-18
; EARLIER APPLICATION NUMBER: 60/055,723
; EARLIER FILING DATE: 1997-08-18
; EARLIER APPLICATION NUMBER: 60/055,948
; EARLIER FILING DATE: 1997-08-18
; EARLIER APPLICATION NUMBER: 60/055,949
; EARLIER FILING DATE: 1997-08-18
; EARLIER APPLICATION NUMBER: 60/055,953
```

```
; EARLIER FILING DATE: 1997-08-18
; EARLIER APPLICATION NUMBER: 60/055,950
; EARLIER FILING DATE: 1997-08-18
; EARLIER APPLICATION NUMBER: 60/055,947
; EARLIER FILING DATE: 1997-08-18
; EARLIER APPLICATION NUMBER: 60/055,964
; EARLIER FILING DATE: 1997-08-18
; EARLIER APPLICATION NUMBER: 60/056,360
; EARLIER FILING DATE: 1997-08-18
; EARLIER APPLICATION NUMBER: 60/055,684
; EARLIER FILING DATE: 1997-08-18
; EARLIER APPLICATION NUMBER: 60/055,984
; EARLIER FILING DATE: 1997-08-18
; EARLIER APPLICATION NUMBER: 60/055,954
; EARLIER FILING DATE: 1997-08-18
; EARLIER APPLICATION NUMBER: 60/058,785
; EARLIER FILING DATE: 1997-09-12
; EARLIER APPLICATION NUMBER: 60/058,664
; EARLIER FILING DATE: 1997-09-12
; EARLIER APPLICATION NUMBER: 60/058,660
; EARLIER FILING DATE: 1997-09-12
; EARLIER APPLICATION NUMBER: 60/058,661
; EARLIER FILING DATE: 1997-09-12
; NUMBER OF SEQ ID NOS: 672
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 589
; LENGTH: 39
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-227-357-589

Query Match      2.5%; Score 6; DB 4; Length 39;
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      59 PWCYVS 64
      |||||
Db      32 PWCYVS 37

RESULT 48
US-09-348-953-6
; Sequence 6, Application US/09348953
; Patent No. 6538103
; GENERAL INFORMATION:
; APPLICANT: Ji, Richard Wei-Dong
; TITLE OF INVENTION: LYSINE BINDING FRAGMENTS OF ANGIOSTATIN
; FILE REFERENCE: DB11Sequences
; CURRENT APPLICATION NUMBER: US/09/348,953
; CURRENT FILING DATE: 1999-07-07
; PRIOR APPLICATION NUMBER: 60/092,831
; PRIOR FILING DATE: 1998-07-14
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 6
; LENGTH: 46
; TYPE: PRT
; ORGANISM: human
US-09-348-953-6

Query Match      2.5%; Score 6; DB 4; Length 46;
Best Local Similarity 100.0%; Pred. No. 1.4e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      48 YCRNPD 53
      |||||
Db      25 YCRNPD 30

RESULT 49
US-09-227-357-588
; Sequence 588, Application US/09227357
```

Patent No. 6342581  
GENERAL INFORMATION:  
APPLICANT: Fischer et al.  
TITLE OF INVENTION: 123 Human Secreted Proteins  
FILE REFERENCE: P201021  
CURRENT APPLICATION NUMBER: US/09/227,357  
CURRENT FILING DATE: 1999-01-08  
EARLIER APPLICATION NUMBER: PCT/US98/13684  
EARLIER FILING DATE: 1998-07-07  
EARLIER APPLICATION NUMBER: 60/051,926  
EARLIER FILING DATE: 1997-07-08  
EARLIER APPLICATION NUMBER: 60/052,793  
EARLIER FILING DATE: 1997-07-08  
EARLIER APPLICATION NUMBER: 60/051,925  
EARLIER FILING DATE: 1997-07-08  
EARLIER APPLICATION NUMBER: 60/051,929  
EARLIER FILING DATE: 1997-07-08  
EARLIER APPLICATION NUMBER: 60/052,803  
EARLIER FILING DATE: 1997-07-08  
EARLIER APPLICATION NUMBER: 60/052,732  
EARLIER FILING DATE: 1997-07-08  
EARLIER APPLICATION NUMBER: 60/051,931  
EARLIER FILING DATE: 1997-07-08  
EARLIER APPLICATION NUMBER: 60/051,932  
EARLIER FILING DATE: 1997-07-08  
EARLIER APPLICATION NUMBER: 60/051,916  
EARLIER FILING DATE: 1997-07-08  
EARLIER APPLICATION NUMBER: 60/051,930  
EARLIER FILING DATE: 1997-07-08  
EARLIER APPLICATION NUMBER: 60/051,918  
EARLIER FILING DATE: 1997-07-08  
EARLIER APPLICATION NUMBER: 60/051,920  
EARLIER FILING DATE: 1997-07-08  
EARLIER APPLICATION NUMBER: 60/052,733  
EARLIER FILING DATE: 1997-07-08  
EARLIER APPLICATION NUMBER: 60/052,795  
EARLIER FILING DATE: 1997-07-08  
EARLIER APPLICATION NUMBER: 60/051,919  
EARLIER FILING DATE: 1997-07-08  
EARLIER APPLICATION NUMBER: 60/051,928  
EARLIER FILING DATE: 1997-07-08  
EARLIER APPLICATION NUMBER: 60/055,722  
EARLIER FILING DATE: 1997-08-18  
EARLIER APPLICATION NUMBER: 60/055,723  
EARLIER FILING DATE: 1997-08-18  
EARLIER APPLICATION NUMBER: 60/055,948  
EARLIER FILING DATE: 1997-08-18  
EARLIER APPLICATION NUMBER: 60/055,949  
EARLIER FILING DATE: 1997-08-18  
EARLIER APPLICATION NUMBER: 60/055,953  
EARLIER FILING DATE: 1997-08-18  
EARLIER APPLICATION NUMBER: 60/055,950  
EARLIER FILING DATE: 1997-08-18  
EARLIER APPLICATION NUMBER: 60/055,947  
EARLIER FILING DATE: 1997-08-18  
EARLIER APPLICATION NUMBER: 60/055,964  
EARLIER FILING DATE: 1997-08-18  
EARLIER APPLICATION NUMBER: 60/056,360  
EARLIER FILING DATE: 1997-08-18  
EARLIER APPLICATION NUMBER: 60/055,684  
EARLIER FILING DATE: 1997-08-18  
EARLIER APPLICATION NUMBER: 60/055,984  
EARLIER FILING DATE: 1997-08-18  
EARLIER APPLICATION NUMBER: 60/055,954  
EARLIER FILING DATE: 1997-08-18  
EARLIER APPLICATION NUMBER: 60/058,785  
EARLIER FILING DATE: 1997-09-12  
EARLIER APPLICATION NUMBER: 60/058,664  
EARLIER FILING DATE: 1997-09-12  
EARLIER APPLICATION NUMBER: 60/058,660  
EARLIER FILING DATE: 1997-09-12  
EARLIER APPLICATION NUMBER: 60/058,661  
EARLIER FILING DATE: 1997-09-12

NUMBER OF SEQ ID NOS: 672  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 588  
LENGTH: 53  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-09-227-357-589  
Query Match 2.5%; Score 6; DB 4; Length 53;  
Best Local Similarity 100.0%; Pred. No. 1.6e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 59 PWCYVS 64  
Db 46 PWCYVS 51  
RESULT 50  
US-08-630-915A-215  
Sequence 215, Application US/08630915A  
Patent No. 6309820  
GENERAL INFORMATION:  
APPLICANT: SPARKS, Andrew B.  
APPLICANT: HOFFMAN, No. 6309820h  
APPLICANT: KAY, Brian K.  
APPLICANT: FOWLES, Dana M.  
APPLICANT: McCONNELL, Stephen J.  
TITLE OF INVENTION: POLYPEPTIDES HAVING A FUNCTIONAL  
TITLE OF INVENTION: DOMAIN OF INTEREST AND METHODS OF IDENTIFYING AND  
TITLE OF INVENTION: USING SAME  
NUMBER OF SEQUENCES: 227  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Pennie & Edmonds LLP  
STREET: 1155 Avenue of the Americas  
CITY: New York  
STATE: New York  
COUNTRY: USA  
ZIP: 10036-2711  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/630,915A  
FILING DATE: 03-APR-1996  
CLASSIFICATION: 536  
ATTORNEY/AGENT INFORMATION:  
NAME: Mirock, S. Leslie  
REGISTRATION NUMBER: 18,872  
REFERENCE/DOCKET NUMBER: 1101-174  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212) 790-9090  
TELEFAX: (212) 869-8864/9741  
TELEX: 66141 PENNIE  
INFORMATION FOR SEQ ID NO: 215:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 59 amino acids  
TYPE: amino acid  
STRANDEDNESS:  
TOPOLOGY: unknown  
MOLECULE TYPE: peptide  
US-08-630-915A-215  
Query Match 2.5%; Score 6; DB 4; Length 59;  
Best Local Similarity 100.0%; Pred. No. 1.7e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 210 VDEKTV 215  
Db 29 VDEKTV 34



RESULT 51  
US-09-489-039A-10500  
; Sequence 10500, Application US/09489039A  
; Patent No. 6610836  
; GENERAL INFORMATION:  
; APPLICANT: Gary Breton et. al  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA  
; FILE REFERENCE: 2709.2004001  
; CURRENT APPLICATION NUMBER: US/09/489,039A  
; CURRENT FILING DATE: 2000-01-27  
; PRIOR APPLICATION NUMBER: US 60/117,747  
; PRIOR FILING DATE: 1999-01-29  
; NUMBER OF SEQ ID NOS: 14342  
; SEQ ID NO 10500  
; LENGTH: 63  
; TYPE: PRT  
; ORGANISM: Klebsiella pneumoniae  
US-09-489-039A-10500  
Query Match 2.5%; Score 6; DB 4; Length 63;  
Best Local Similarity 100.0%; Pred. No. 1.8e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 85 SQUALPA 90  
Db 46 SQUALPA 51  
RESULT 52  
US-09-107-532A-6727  
; Sequence 6727, Application US/09107532A  
; Patent No. 6583275  
; GENERAL INFORMATION:  
; APPLICANT: Lynn A Doucette-Stamm and David Bush  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO  
; ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS  
; NUMBER OF SEQUENCES: 7310  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: GENOME THERAPEUTICS CORPORATION  
; STREET: 100 Beaver Street  
; CITY: Waltham  
; STATE: Massachusetts  
; COUNTRY: USA  
; ZIP: 02354  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: CD-ROM ISO9660  
; COMPUTER: PC  
; OPERATING SYSTEM: <Unknown>  
; SOFTWARE: ASCII  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/107,532A  
; FILING DATE: 30-Jun-1998  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 60/085,598  
; FILING DATE: 14 May 1998  
; APPLICATION NUMBER: 60/051571  
; FILING DATE: July 2, 1997  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Ariniello, Pamela Deneke  
; REGISTRATION NUMBER: 40,489  
; REFERENCE/DOCKET NUMBER: GTC-012  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (781)893-5007  
; TELEFAX: (781)893-8277  
; INFORMATION FOR SEQ ID NO: 6727:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 66 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; HYPOTHETICAL: YES  
; ORIGINAL SOURCE:

; ORGANISM: Enterococcus faecium  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (B) LOCATION 1...66  
; SEQUENCE DESCRIPTION: SEQ ID NO: 6727:  
US-09-107-532A-6727  
Query Match 2.5%; Score 6; DB 4; Length 66;  
Best Local Similarity 100.0%; Pred. No. 1.9e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 162 AIGAGI 167  
Db 45 AIGAGI 50  
RESULT 53  
US-09-134-000C-4771  
; Sequence 4771, Application US/09134000C  
; Patent No. 6617156  
; GENERAL INFORMATION:  
; APPLICANT: Lynn Doucette-Stamm et al  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO  
; ENTEROCOCCUS FAECALIS FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: 032796-032  
; CURRENT APPLICATION NUMBER: US/09/134,000C  
; CURRENT FILING DATE: 1998-08-13  
; PRIOR APPLICATION NUMBER: US 60/055,778  
; PRIOR FILING DATE: 1997-08-15  
; NUMBER OF SEQ ID NOS: 6812  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 4771  
; LENGTH: 67  
; TYPE: PRT  
; ORGANISM: Enterococcus faecalis  
US-09-134-000C-4771  
Query Match 2.5%; Score 6; DB 4; Length 67;  
Best Local Similarity 100.0%; Pred. No. 1.9e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 33 SGLASA 38  
Db 21 SGLASA 26  
RESULT 54  
US-09-489-039A-9351  
; Sequence 9351, Application US/09489039A  
; Patent No. 6610936  
; GENERAL INFORMATION:  
; APPLICANT: Gary Breton et. al  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA  
; FILE REFERENCE: 2709.2004001  
; CURRENT APPLICATION NUMBER: US/09/489,039A  
; CURRENT FILING DATE: 2000-01-27  
; PRIOR APPLICATION NUMBER: US 60/117,747  
; PRIOR FILING DATE: 1999-01-29  
; NUMBER OF SEQ ID NOS: 14342  
; SEQ ID NO 9351  
; LENGTH: 72  
; TYPE: PRT  
; ORGANISM: Klebsiella pneumoniae  
US-09-489-039A-9351  
Query Match 2.5%; Score 6; DB 4; Length 72;  
Best Local Similarity 100.0%; Pred. No. 2.1e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 36 ASAPVS 41  
Db 52 ASAPVS 57

us-10-057-951-2\_copy\_22\_263.oligo.ra

RESULT 55

US-09-107-532A-4213

Sequence 4213, Application US/09107532A

Patent No. 6583275

GENERAL INFORMATION:

APPLICANT: Lynn A Doucette-Stamm and David Bush

TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO

TITLE OF INVENTION: ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS

NUMBER OF SEQUENCES: 7310

CORRESPONDENCE ADDRESS:

ADDRESSEE: GENOME THERAPEUTICS CORPORATION

STREET: 100 Beaver Street

CITY: Waltham

STATE: Massachusetts

COUNTRY: USA

ZIP: 02354

COMPUTER READABLE FORM:

MEDIUM TYPE: CD-ROM ISO9660

COMPUTER: PC

OPERATING SYSTEM: <Unknown>

SOFTWARE: ASCII

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/107,532A

FILING DATE: 30-Jun-1998

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 60/085,598

FILING DATE: 14 May 1998

APPLICATION NUMBER: 60/051571

FILING DATE: July 2, 1997

ATTORNEY/AGENT INFORMATION:

NAME: Ariniello, Pamela Deneke

REGISTRATION NUMBER: 40,489

REFERENCE/DOCKET NUMBER: GTC-012

TELECOMMUNICATION INFORMATION:

TELEPHONE: (781)893-5007

TELEFAX: (781)893-8277

INFORMATION FOR SEQ ID NO: 4213:

SEQUENCE CHARACTERISTICS:

LENGTH: 73 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

HYPOTHETICAL: YES

ORIGINAL SOURCE:

ORGANISM: Enterococcus faecium

FEATURE:

NAME/KEY: misc.feature

LOCATION: (B) LOCATION 1...73

SEQUENCE DESCRIPTION: SEQ ID NO: 4213:

US-09-107-532A-4213

Query Match 2.5%; Score 6; DB 4; Length 73;

Best Local Similarity 100.0%; Pred. No. 2.1e+02;

Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 66 EAGVPE 71

DB 40 EAGVPE 45

RESULT 56

US-08-612-788-13

Sequence 13, Application US/08612788

Patent No. 5837682

GENERAL INFORMATION:

APPLICANT: Folkman, M. Judah

APPLICANT: O'Reilly, Micheal

APPLICANT: Cao, Yihai

APPLICANT: Sim, B. Kim Lee

TITLE OF INVENTION: Angiostatin Fragments and Method of Use

NUMBER OF SEQUENCES: 45

CORRESPONDENCE ADDRESS:

ADDRESSEE: Jones & Askew

STREET: 191 Peachtree Street, 37th Floor

CITY: Atlanta

STATE: Georgia

COUNTRY: U.S.

ZIP: 30303-1769

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent in Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/612,788

QY 48 YCRNPD 53

DB 49 YCRNPD 54

RESULT 57

US-08-612-788-14

Sequence 14, Application US/08612788

Patent No. 5837682

GENERAL INFORMATION:

APPLICANT: Folkman, M. Judah

APPLICANT: O'Reilly, Micheal

APPLICANT: Cao, Yihai

APPLICANT: Sim, B. Kim Lee

TITLE OF INVENTION: Angiostatin Fragments and Method of Use

NUMBER OF SEQUENCES: 45

CORRESPONDENCE ADDRESS:

ADDRESSEE: Jones & Askew

STREET: 191 Peachtree Street, 37th Floor

CITY: Atlanta

STATE: Georgia

COUNTRY: U.S.

ZIP: 30303-1769

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent in Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/612,788

QY 48 YCRNPD 53

DB 49 YCRNPD 54

us-10-057-951-2\_copy\_22\_263.oligo.ra

RESULT 55

US-09-107-532A-4213

Sequence 4213, Application US/09107532A

Patent No. 6583275

GENERAL INFORMATION:

APPLICANT: Lynn A Doucette-Stamm and David Bush

TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO

TITLE OF INVENTION: ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS

NUMBER OF SEQUENCES: 7310

CORRESPONDENCE ADDRESS:

ADDRESSEE: GENOME THERAPEUTICS CORPORATION

STREET: 100 Beaver Street

CITY: Waltham

STATE: Massachusetts

COUNTRY: USA

ZIP: 02354

COMPUTER READABLE FORM:

MEDIUM TYPE: CD-ROM ISO9660

COMPUTER: PC

OPERATING SYSTEM: <Unknown>

SOFTWARE: ASCII

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/107,532A

FILING DATE: 30-Jun-1998

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 60/085,598

FILING DATE: 14 May 1998

APPLICATION NUMBER: 60/051571

FILING DATE: July 2, 1997

ATTORNEY/AGENT INFORMATION:

NAME: Ariniello, Pamela Deneke

REGISTRATION NUMBER: 40,489

REFERENCE/DOCKET NUMBER: GTC-012

TELECOMMUNICATION INFORMATION:

TELEPHONE: (781)893-5007

TELEFAX: (781)893-8277

INFORMATION FOR SEQ ID NO: 4213:

SEQUENCE CHARACTERISTICS:

LENGTH: 73 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

HYPOTHETICAL: YES

ORIGINAL SOURCE:

ORGANISM: Enterococcus faecium

FEATURE:

NAME/KEY: misc.feature

LOCATION: (B) LOCATION 1...73

SEQUENCE DESCRIPTION: SEQ ID NO: 4213:

US-09-107-532A-4213

Query Match 2.5%; Score 6; DB 4; Length 73;

Best Local Similarity 100.0%; Pred. No. 2.1e+02;

Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 66 EAGVPE 71

DB 40 EAGVPE 45

RESULT 56

US-08-612-788-13

Sequence 13, Application US/08612788

Patent No. 5837682

GENERAL INFORMATION:

APPLICANT: Folkman, M. Judah

APPLICANT: O'Reilly, Micheal

APPLICANT: Cao, Yihai

APPLICANT: Sim, B. Kim Lee

TITLE OF INVENTION: Angiostatin Fragments and Method of Use

NUMBER OF SEQUENCES: 45

CORRESPONDENCE ADDRESS:

ADDRESSEE: Jones & Askew

STREET: 191 Peachtree Street, 37th Floor

CITY: Atlanta

STATE: Georgia

COUNTRY: U.S.

ZIP: 30303-1769

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent in Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/612,788

QY 48 YCRNPD 53

DB 49 YCRNPD 54

RESULT 57

US-08-612-788-14

Sequence 14, Application US/08612788

Patent No. 5837682

GENERAL INFORMATION:

APPLICANT: Folkman, M. Judah

APPLICANT: O'Reilly, Micheal

APPLICANT: Cao, Yihai

APPLICANT: Sim, B. Kim Lee

TITLE OF INVENTION: Angiostatin Fragments and Method of Use

NUMBER OF SEQUENCES: 45

CORRESPONDENCE ADDRESS:

ADDRESSEE: Jones & Askew

STREET: 191 Peachtree Street, 37th Floor

CITY: Atlanta

STATE: Georgia

COUNTRY: U.S.

ZIP: 30303-1769

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent in Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/612,788

QY 48 YCRNPD 53

DB 49 YCRNPD 54



```
QY      48 YCRNPD 53
Db      49 YCRNPD 54

RESULT 60
US-08-612-788-17
; Sequence 17, Application US/08612788
; Patent No. 5837682
; GENERAL INFORMATION:
; APPLICANT: Folkman, M. Judah
; APPLICANT: O'Reilly, Micheal
; APPLICANT: Cao, Yihai
; APPLICANT: Sim, B. Kim Lee
; TITLE OF INVENTION: Angiostatin Fragments and Method of Use
; NUMBER OF SEQUENCES: 45
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Jones & Askew
; STREET: 191 Peachtree Street, 37th Floor
; CITY: Atlanta
; STATE: Georgia
; COUNTRY: U.S.
; ZIP: 30303-1769
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/612,788
; FILING DATE:
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Warren, William L.
; REGISTRATION NUMBER: 36,714
; REFERENCE/DOCKET NUMBER: 05213-0126
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 404-818-3700
; TELEFAX: 404-818-3799
; INFORMATION FOR SEQ ID NO: 17:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 78 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; FRAGMENT TYPE: N-terminal
; ORIGINAL SOURCE:
; ORGANISM: Murine
; IMMEDIATE SOURCE:
; CLONE: K3
US-08-612-788-17

Query Match      2.5%; Score 6; DB 2; Length 78;
Best Local Similarity 100.0%; Pred. No. 2.2e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      48 YCRNPD 53
Db      49 YCRNPD 54

RESULT 61
US-08-612-788-18
; Sequence 18, Application US/08612788
; Patent No. 5837682
; GENERAL INFORMATION:
; APPLICANT: Folkman, M. Judah
; APPLICANT: O'Reilly, Micheal
; APPLICANT: Cao, Yihai
; TITLE OF INVENTION: Angiostatin Fragments and Method of Use
; NUMBER OF SEQUENCES: 45
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Jones & Askew
; STREET: 191 Peachtree Street, 37th Floor
; CITY: Atlanta
; STATE: Georgia
; COUNTRY: U.S.
; ZIP: 30303-1769
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/612,788
; FILING DATE:
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Warren, William L.
; REGISTRATION NUMBER: 36,714
; REFERENCE/DOCKET NUMBER: 05213-0126
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 404-818-3700
; TELEFAX: 404-818-3799
; INFORMATION FOR SEQ ID NO: 18:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 78 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; FRAGMENT TYPE: N-terminal
; ORIGINAL SOURCE:
; ORGANISM: Homo sapiens
; IMMEDIATE SOURCE:
; CLONE: K3
US-08-612-788-18

Query Match      2.5%; Score 6; DB 2; Length 78;
Best Local Similarity 100.0%; Pred. No. 2.2e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      48 YCRNPD 53
Db      49 YCRNPD 54

RESULT 62
US-08-612-788-19
; Sequence 19, Application US/08612788
; Patent No. 5837682
; GENERAL INFORMATION:
; APPLICANT: Folkman, M. Judah
; APPLICANT: O'Reilly, Micheal
; APPLICANT: Cao, Yihai
; APPLICANT: Sim, B. Kim Lee
; TITLE OF INVENTION: Angiostatin Fragments and Method of Use
; NUMBER OF SEQUENCES: 45
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Jones & Askew
; STREET: 191 Peachtree Street, 37th Floor
; CITY: Atlanta
; STATE: Georgia
; COUNTRY: U.S.
; ZIP: 30303-1769
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/612,788
; FILING DATE:
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Warren, William L.
; REGISTRATION NUMBER: 36,714
; REFERENCE/DOCKET NUMBER: 05213-0126
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 404-818-3700
; TELEFAX: 404-818-3799
; INFORMATION FOR SEQ ID NO: 18:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 78 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; FRAGMENT TYPE: N-terminal
; ORIGINAL SOURCE:
; ORGANISM: Homo sapiens
; IMMEDIATE SOURCE:
; CLONE: K3
US-08-612-788-19

Query Match      2.5%; Score 6; DB 2; Length 78;
Best Local Similarity 100.0%; Pred. No. 2.2e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      48 YCRNPD 53
Db      49 YCRNPD 54
```

SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/612,788  
FILING DATE:  
CLASSIFICATION: 514  
ATTORNEY/AGENT INFORMATION:  
NAME: Warren, William L.  
REGISTRATION NUMBER: 36,714  
REFERENCE/DOCKET NUMBER: 05213-0126  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 404-818-3700  
TELEFAX: 404-818-3799  
INFORMATION FOR SEQ ID NO: 19:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 78 amino acids  
TYPE: amino acid  
STRANDEDNESS:  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
FRAGMENT TYPE: N-terminal  
ORIGINAL SOURCE: Rhesus monkey  
ORGANISM:  
IMMEDIATE SOURCE:  
CLONE: K3  
US-08-612-788-19

Query Match 2.5%; Score 6; DB 2; Length 78;  
Best Local Similarity 100.0%; Pred. No. 2.2e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0;

QY 48 YCRNPD 53  
DB 49 YCRNPD 54

RESULT 63  
US-08-612-788-20  
Sequence 20, Application US/08612788  
Patent No. 5837682  
GENERAL INFORMATION:  
APPLICANT: Folkman, M. Judah  
APPLICANT: O'Reilly, Micheal  
APPLICANT: Cao, Yihai  
APPLICANT: Sim, B. Kim Lee  
TITLE OF INVENTION: Angiostatin Fragments and Method of Use  
NUMBER OF SEQUENCES: 45  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Jones & Askew  
STREET: 191 Peachtree Street, 37th Floor  
CITY: Atlanta  
STATE: Georgia  
COUNTRY: U.S.  
ZIP: 30303-1769  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/612,788  
FILING DATE:  
CLASSIFICATION: 514  
ATTORNEY/AGENT INFORMATION:  
NAME: Warren, William L.  
REGISTRATION NUMBER: 36,714  
REFERENCE/DOCKET NUMBER: 05213-0126  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 404-818-3700  
TELEFAX: 404-818-3799  
INFORMATION FOR SEQ ID NO: 20:  
SEQUENCE CHARACTERISTICS:

LENGTH: 78 amino acids  
TYPE: amino acid  
STRANDEDNESS:  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
FRAGMENT TYPE: N-terminal  
ORIGINAL SOURCE: Porcine  
ORGANISM:  
IMMEDIATE SOURCE:  
CLONE: K3  
US-08-612-788-20

Query Match 2.5%; Score 6; DB 2; Length 78;  
Best Local Similarity 100.0%; Pred. No. 2.2e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0;

QY 48 YCRNPD 53  
DB 49 YCRNPD 54

RESULT 64  
US-08-612-788-21  
Sequence 21, Application US/08612788  
Patent No. 5837682  
GENERAL INFORMATION:  
APPLICANT: Folkman, M. Judah  
APPLICANT: O'Reilly, Micheal  
APPLICANT: Cao, Yihai  
APPLICANT: Sim, B. Kim Lee  
TITLE OF INVENTION: Angiostatin Fragments and Method of Use  
NUMBER OF SEQUENCES: 45  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Jones & Askew  
STREET: 191 Peachtree Street, 37th Floor  
CITY: Atlanta  
STATE: Georgia  
COUNTRY: U.S.  
ZIP: 30303-1769  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/612,788  
FILING DATE:  
CLASSIFICATION: 514  
ATTORNEY/AGENT INFORMATION:  
NAME: Warren, William L.  
REGISTRATION NUMBER: 36,714  
REFERENCE/DOCKET NUMBER: 05213-0126  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 404-818-3700  
TELEFAX: 404-818-3799  
INFORMATION FOR SEQ ID NO: 21:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 78 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
FRAGMENT TYPE: N-terminal  
ORIGINAL SOURCE: Bovine  
ORGANISM:  
IMMEDIATE SOURCE:  
CLONE: K3  
US-08-612-788-21

Query Match 2.5%; Score 6; DB 2; Length 78;  
Best Local Similarity 100.0%; Pred. No. 2.2e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 48 YCRNPD 53  
Db 49 YCRNPD 54

RESULT 65  
US-08-612-788-22  
; Sequence 22, Application US/08612788  
; Patent No. 5837682  
; GENERAL INFORMATION:  
; APPLICANT: Folkman, M. Judah  
; APPLICANT: O'Reilly, Micheal  
; APPLICANT: Cao, Yihai  
; APPLICANT: Sim, B. Kim Lee  
; TITLE OF INVENTION: Angiostatin Fragments and Method of Use  
; NUMBER OF SEQUENCES: 45  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Jones & Askew  
; STREET: 191 Peachtree Street, 37th Floor  
; CITY: Atlanta  
; STATE: Georgia  
; COUNTRY: U.S.  
; ZIP: 30303-1769  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/612,788  
; FILING DATE:  
; CLASSIFICATION: 514  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Warren, William L.  
; REGISTRATION NUMBER: 36,714  
; REFERENCE/DOCKET NUMBER: 05213-0126  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 404-818-3700  
; TELEFAX: 404-818-3799  
; INFORMATION FOR SEQ ID NO: 22:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 78 amino acids  
; TYPE: amino acid  
; STRANDEDNESS:  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; HYPOTHETICAL: NO  
; ANTI-SENSE: NO  
; FRAGMENT TYPE: N-terminal  
; ORIGINAL SOURCE:  
; ORGANISM: Murine  
; IMMEDIATE SOURCE:  
; CLONE: K4  
US-08-612-788-22

Query Match 2.5%; Score 6; DB 2; Length 78;  
Best Local Similarity 100.0%; Pred. No. 2.2e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 48 YCRNPD 53  
Db 49 YCRNPD 54

RESULT 66  
US-08-612-788-23  
; Sequence 23, Application US/08612788  
; Patent No. 5837682  
; GENERAL INFORMATION:

APPLICANT: Folkman, M. Judah  
APPLICANT: O'Reilly, Micheal  
APPLICANT: Cao, Yihai  
APPLICANT: Sim, B. Kim Lee  
TITLE OF INVENTION: Angiostatin Fragments and Method of Use  
NUMBER OF SEQUENCES: 45  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Jones & Askew  
STREET: 191 Peachtree Street, 37th Floor  
CITY: Atlanta  
STATE: Georgia  
COUNTRY: U.S.  
ZIP: 30303-1769  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/612,788  
FILING DATE:  
CLASSIFICATION: 514  
ATTORNEY/AGENT INFORMATION:  
NAME: Warren, William L.  
REGISTRATION NUMBER: 36,714  
REFERENCE/DOCKET NUMBER: 05213-0126  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 404-818-3700  
TELEFAX: 404-818-3799  
INFORMATION FOR SEQ ID NO: 23:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 78 amino acids  
TYPE: amino acid  
STRANDEDNESS:  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
FRAGMENT TYPE: N-terminal  
ORIGINAL SOURCE:  
ORGANISM: Homo sapiens  
IMMEDIATE SOURCE:  
CLONE: K4  
US-08-612-788-23

Query Match 2.5%; Score 6; DB 2; Length 78;  
Best Local Similarity 100.0%; Pred. No. 2.2e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 48 YCRNPD 53  
Db 49 YCRNPD 54

RESULT 67  
US-08-763-528A-3  
; Sequence 3, Application US/08763528A  
; Patent No. 5854221  
; GENERAL INFORMATION:  
; APPLICANT: Cao, Yihai  
; APPLICANT: Folkman, M. Judah  
; TITLE OF INVENTION: Endothelial Cell Proliferation Inhibitor  
; TITLE OF INVENTION: and Method of Use  
; NUMBER OF SEQUENCES: 6  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Jones & Askew, LLP  
; STREET: 191 Peachtree Street, 37th Floor  
; CITY: Atlanta  
; STATE: Georgia  
; COUNTRY: US  
; ZIP: 30303  
; COMPUTER READABLE FORM: disk  
; MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/763,528A  
FILING DATE: 12-DEC-1996  
CLASSIFICATION: 530  
ATTORNEY/AGENT INFORMATION:  
NAME: Warren, William L.  
REGISTRATION NUMBER: 36,714  
REFERENCE/DOCKET NUMBER: 05940-0251  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 404-818-3700  
TELEFAX: 404-818-3799  
INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 78 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
MOLECULE TYPE: protein  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
FRAGMENT TYPE: internal  
FEATURE:  
NAME/KEY: Protein  
LOCATION: 1..78  
OTHER INFORMATION: /note= "Kringle 2 - Figure 3"  
US-08-763-528A-3

Query Match 2.5%; Score 6; DB 2; Length 78;  
Best Local Similarity 100.0%; Pred. No. 2.2e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 48 YCRNPD 53  
DB 49 YCRNPD 54

RESULT 68  
US-08-763-528A-4  
Sequence 4, Application US/08763528A  
Patent No. 5854221  
GENERAL INFORMATION:  
APPLICANT: Cao, Yihai  
APPLICANT: Folkman, M. Judah  
TITLE OF INVENTION: Endothelial Cell Proliferation Inhibitor  
TITLE OF INVENTION: and Method of Use  
NUMBER OF SEQUENCES: 6  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Jones & Askew, LLP  
STREET: 191 Peachtree Street, 37th Floor  
CITY: Atlanta  
STATE: Georgia  
COUNTRY: US  
ZIP: 30303  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/763,528A  
FILING DATE: 12-DEC-1996  
CLASSIFICATION: 530  
ATTORNEY/AGENT INFORMATION:  
NAME: Warren, William L.  
REGISTRATION NUMBER: 36,714  
REFERENCE/DOCKET NUMBER: 05940-0251  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 404-818-3700  
TELEFAX: 404-818-3799  
INFORMATION FOR SEQ ID NO: 4:

SEQUENCE CHARACTERISTICS:  
LENGTH: 78 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
FRAGMENT TYPE: internal  
FEATURE:  
NAME/KEY: Protein  
LOCATION: 1..78  
OTHER INFORMATION: /note= "Kringle 3 - Figure 3"  
US-08-763-528A-4

Query Match 2.5%; Score 6; DB 2; Length 78;  
Best Local Similarity 100.0%; Pred. No. 2.2e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 48 YCRNPD 53  
DB 49 YCRNPD 54

RESULT 69  
US-08-763-528A-5  
Sequence 5, Application US/08763528A  
Patent No. 5854221  
GENERAL INFORMATION:  
APPLICANT: Cao, Yihai  
APPLICANT: Folkman, M. Judah  
TITLE OF INVENTION: Endothelial Cell Proliferation Inhibitor  
TITLE OF INVENTION: and Method of Use  
NUMBER OF SEQUENCES: 6  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Jones & Askew, LLP  
STREET: 191 Peachtree Street, 37th Floor  
CITY: Atlanta  
STATE: Georgia  
COUNTRY: US  
ZIP: 30303  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/763,528A  
FILING DATE: 12-DEC-1996  
CLASSIFICATION: 530  
ATTORNEY/AGENT INFORMATION:  
NAME: Warren, William L.  
REGISTRATION NUMBER: 36,714  
REFERENCE/DOCKET NUMBER: 05940-0251  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 404-818-3700  
TELEFAX: 404-818-3799  
INFORMATION FOR SEQ ID NO: 5:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 78 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
FRAGMENT TYPE: internal  
FEATURE:  
NAME/KEY: Protein  
LOCATION: 1..78  
OTHER INFORMATION: /note= "Kringle 4 - Figure 3"  
US-08-763-528A-5

Query Match 2.5%; Score 6; DB 2; Length 78;  
Best Local Similarity 100.0%; Pred. No. 2.2e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 48 YCRNPD 53  
| | | | |  
Db 49 YCRNPD 54

RESULT 70  
US-09-066-028-13  
; Sequence 13, Application US/09066028  
; Patent No. 6024688  
; GENERAL INFORMATION:  
; APPLICANT: Folkman, M. Judah  
; APPLICANT: O'Reilly, Micheal  
; APPLICANT: Cao, Yihai  
; APPLICANT: Sim, B. Kim Lee  
; TITLE OF INVENTION: Angiostatin Fragments and Method of Use  
; NUMBER OF SEQUENCES: 45  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Jones & Askew  
; STREET: 191 Peachtree Street, 37th Floor  
; CITY: Atlanta  
; STATE: Georgia  
; COUNTRY: U.S.  
; ZIP: 30303-1769  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent in Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/066,028  
; FILING DATE:  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA: 08/612,788  
; APPLICATION NUMBER:  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Warren, William L.  
; REGISTRATION NUMBER: 36,714  
; REFERENCE/DOCKET NUMBER: 05213-0126  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 404-818-3700  
; TELEFAX: 404-818-3799  
; INFORMATION FOR SEQ ID NO: 13:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 78 amino acids  
; TYPE: amino acid  
; STRANDEDNESS:  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; HYPOTHETICAL: NO  
; ANTI-SENSE: NO  
; FRAGMENT TYPE: N-terminal  
; ORIGINAL SOURCE:  
; ORGANISM: Homo sapien  
; IMMEDIATE SOURCE:  
; CLONE: K2  
; US-09-066-028-13

Query Match 2.5%; Score 6; DB 3; Length 78;  
Best Local Similarity 100.0%; Pred. No. 2.2e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 48 YCRNPD 53  
| | | | |  
Db 49 YCRNPD 54

RESULT 71  
US-09-066-028-14  
; Sequence 14, Application US/09066028  
; Patent No. 6024688  
; GENERAL INFORMATION:  
; APPLICANT: Folkman, M. Judah  
; APPLICANT: O'Reilly, Micheal  
; APPLICANT: Cao, Yihai  
; APPLICANT: Sim, B. Kim Lee  
; TITLE OF INVENTION: Angiostatin Fragments and Method of Use  
; NUMBER OF SEQUENCES: 45  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Jones & Askew  
; STREET: 191 Peachtree Street, 37th Floor  
; CITY: Atlanta  
; STATE: Georgia  
; COUNTRY: U.S.  
; ZIP: 30303-1769  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent in Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/066,028  
; FILING DATE:  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA: 08/612,788  
; APPLICATION NUMBER:  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Warren, William L.  
; REGISTRATION NUMBER: 36,714  
; REFERENCE/DOCKET NUMBER: 05213-0126  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 404-818-3700  
; TELEFAX: 404-818-3799  
; INFORMATION FOR SEQ ID NO: 14:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 78 amino acids  
; TYPE: amino acid  
; STRANDEDNESS:  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; HYPOTHETICAL: NO  
; ANTI-SENSE: NO  
; FRAGMENT TYPE: N-terminal  
; ORIGINAL SOURCE:  
; ORGANISM: Homo sapien  
; IMMEDIATE SOURCE:  
; CLONE: K2  
; US-09-066-028-14

Query Match 2.5%; Score 6; DB 3; Length 78;  
Best Local Similarity 100.0%; Pred. No. 2.2e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 48 YCRNPD 53  
| | | | |  
Db 49 YCRNPD 54

RESULT 72  
US-09-066-028-15

Query Match 2.5%; Score 6; DB 3; Length 78;  
Best Local Similarity 100.0%; Pred. No. 2.2e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 48 YCRNPD 53  
| | | | |  
Db 49 YCRNPD 54

RESULT 72  
US-09-066-028-15  
; Sequence 15, Application US/09066028  
; Patent No. 6024688  
; GENERAL INFORMATION:  
; APPLICANT: Folkman, M. Judah  
; APPLICANT: O'Reilly, Micheal  
; APPLICANT: Cao, Yihai  
; APPLICANT: Sim, B. Kim Lee  
; TITLE OF INVENTION: Angiostatin Fragments and Method of Use  
; NUMBER OF SEQUENCES: 45  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Jones & Askew  
; STREET: 191 Peachtree Street, 37th Floor  
; CITY: Atlanta  
; STATE: Georgia  
; COUNTRY: U.S.  
; ZIP: 30303-1769  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent in Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/066,028  
; FILING DATE:  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA: 08/612,788  
; APPLICATION NUMBER:  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Warren, William L.  
; REGISTRATION NUMBER: 36,714  
; REFERENCE/DOCKET NUMBER: 05213-0126  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 404-818-3700  
; TELEFAX: 404-818-3799  
; INFORMATION FOR SEQ ID NO: 14:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 78 amino acids  
; TYPE: amino acid  
; STRANDEDNESS:  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; HYPOTHETICAL: NO  
; ANTI-SENSE: NO  
; FRAGMENT TYPE: N-terminal  
; ORIGINAL SOURCE:  
; ORGANISM: Rhesus monkey  
; IMMEDIATE SOURCE:  
; CLONE: K2  
; US-09-066-028-14

Query Match 2.5%; Score 6; DB 3; Length 78;  
Best Local Similarity 100.0%; Pred. No. 2.2e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 48 YCRNPD 53  
| | | | |  
Db 49 YCRNPD 54

RESULT 72  
US-09-066-028-15  
; Sequence 15, Application US/09066028  
; Patent No. 6024688  
; GENERAL INFORMATION:  
; APPLICANT: Folkman, M. Judah  
; APPLICANT: O'Reilly, Micheal  
; APPLICANT: Cao, Yihai  
; APPLICANT: Sim, B. Kim Lee  
; TITLE OF INVENTION: Angiostatin Fragments and Method of Use  
; NUMBER OF SEQUENCES: 45  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Jones & Askew



STREET: 191 Peachtree Street, 37th Floor  
CITY: Atlanta  
STATE: Georgia  
COUNTRY: U.S.  
ZIP: 30303-1769  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/066,028  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/612,788  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Warren, William L.  
REGISTRATION NUMBER: 36,714  
REFERENCE/DOCKET NUMBER: 05213-0126  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 404-818-3700  
TELEFAX: 404-818-3799  
INFORMATION FOR SEQ ID NO: 15:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 78 amino acids  
TYPE: amino acid  
STRANDEDNESS:  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
FRAGMENT TYPE: N-terminal  
ORIGINAL SOURCE:  
ORGANISM: Porcine  
IMMEDIATE SOURCE:  
CLONE: K2  
US-09-066-028-15

Query Match 2.5%; Score 6; DB 3; Length 78;  
Best Local Similarity 100.0%; Pred. No. 2.2e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 48 YCRNPD 53  
Db 49 YCRNPD 54

RESULT 73  
US-09-066-028-16  
Sequence 16, Application US/09066028  
Patent No. 6024688  
GENERAL INFORMATION:  
APPLICANT: Folkman, M. Judah  
APPLICANT: O'Reilly, Micheal  
APPLICANT: Cao, Yihai  
APPLICANT: Sim, B. Kim Lee  
TITLE OF INVENTION: Angiostatin Fragments and Method of Use  
NUMBER OF SEQUENCES: 45  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Jones & Askew  
STREET: 191 Peachtree Street, 37th Floor  
CITY: Atlanta  
STATE: Georgia  
COUNTRY: U.S.  
ZIP: 30303-1769  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/066,028  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/612,788  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Warren, William L.  
REGISTRATION NUMBER: 36,714  
REFERENCE/DOCKET NUMBER: 05213-0126  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 404-818-3700  
TELEFAX: 404-818-3799  
INFORMATION FOR SEQ ID NO: 16:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 78 amino acids  
TYPE: amino acid  
STRANDEDNESS:  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
FRAGMENT TYPE: N-terminal  
ORIGINAL SOURCE:  
ORGANISM: Bovine  
IMMEDIATE SOURCE:  
CLONE: K2  
US-09-066-028-16  
Query Match 2.5%; Score 6; DB 3; Length 78;  
Best Local Similarity 100.0%; Pred. No. 2.2e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 48 YCRNPD 53  
Db 49 YCRNPD 54

RESULT 74  
US-09-066-028-17  
Sequence 17, Application US/09066028  
Patent No. 6024688  
GENERAL INFORMATION:  
APPLICANT: Folkman, M. Judah  
APPLICANT: O'Reilly, Micheal  
APPLICANT: Cao, Yihai  
APPLICANT: Sim, B. Kim Lee  
TITLE OF INVENTION: Angiostatin Fragments and Method of Use  
NUMBER OF SEQUENCES: 45  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Jones & Askew  
STREET: 191 Peachtree Street, 37th Floor  
CITY: Atlanta  
STATE: Georgia  
COUNTRY: U.S.  
ZIP: 30303-1769  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/066,028  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/612,788  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Warren, William L.  
REGISTRATION NUMBER: 36,714  
REFERENCE/DOCKET NUMBER: 05213-0126  
TELECOMMUNICATION INFORMATION:

TELEPHONE: 404-818-3700  
TELEFAX: 404-818-3799  
INFORMATION FOR SEQ ID NO: 17:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 78 amino acids  
TYPE: amino acid  
STRANDEDNESS:  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
HYPOTHETICAL: NO  
ANTI-SENSE: NO N-terminal  
FRAGMENT TYPE:  
ORIGINAL SOURCE:  
ORGANISM: Murine  
IMMEDIATE SOURCE:  
CLONE: K3  
US-09-066-028-17

Query Match 2.5%; Score 6; DB 3; Length 78;  
Best Local Similarity 100.0%; Pred. No. 2.2e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 48 YCRNPD 53  
Db 49 YCRNPD 54

RESULT 75  
US-09-066-028-18  
Sequence 18, Application US/09066028  
Patent No. 6024688  
GENERAL INFORMATION:  
APPLICANT: Folkman, M. Judah  
APPLICANT: O'Reilly, Micheal  
APPLICANT: Cao, Yihai  
APPLICANT: Sim, B. Kim Lee  
TITLE OF INVENTION: Angiostatin Fragments and Method of Use  
NUMBER OF SEQUENCES: 45  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Jones & Askew  
STREET: 191 Peachtree Street, 37th Floor  
CITY: Atlanta  
STATE: Georgia  
COUNTRY: U.S.  
ZIP: 30303-1769  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/066,028  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/612,788  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Warren, William L.  
REGISTRATION NUMBER: 36,714  
REFERENCE/DOCKET NUMBER: 05213-0126  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 404-818-3700  
TELEFAX: 404-818-3799  
INFORMATION FOR SEQ ID NO: 18:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 78 amino acids  
TYPE: amino acid  
STRANDEDNESS:  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
HYPOTHETICAL: NO  
ANTI-SENSE: NO

FRAGMENT TYPE: N-terminal  
ORIGINAL SOURCE:  
ORGANISM: Homo sapiens  
IMMEDIATE SOURCE:  
CLONE: K3  
US-09-066-028-18

Query Match 2.5%; Score 6; DB 3; Length 78;  
Best Local Similarity 100.0%; Pred. No. 2.2e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 48 YCRNPD 53  
Db 49 YCRNPD 54

Search completed: March 17, 2004, 07:27:22  
Job time : 24 secs